Towards Defining the Molecular Mechanism of Hygienic Behaviour in Honey Bees (*Apis mellifera*)

by

Alison McAfee

B.Sc., The University of British Columbia, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Genome Science and Technology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2018

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Towards Defining the Molecular Mechanism of Hygienic Behaviour in Honey Bees (*Apis mellifera*)

---

submitted by	Alison McAfee	in partial fulfillment of the requirements
for the degree of	Doctor of Philosophy	
in	The Faculty of Graduate and Postdoctoral Studies	

Examining Committee:

Dr. Leonard Foster

Supervisor

Dr. Christopher Overall

Supervisory Committee Member

Supervisory Committee Member

Dr. Patricia Schulte

University Examiner

Dr. Murray Isman

University Examiner

Additional Supervisory Committee Members:

Dr. Erika Plettner

Supervisory Committee Member

Dr. Michael Gordon

Supervisory Committee Member

Abstract

Honey bees (*Apis mellifera*) are integral components of the agricultural industry, but diseases and parasites like the *Varroa destructor* mite threaten their health and longevity. Some honey bee colonies harbor natural disease-resistance traits, and proteomics has been a fruitful tool to investigate mechanisms of disease resistance; however, *Varroa* proteomics is a budding field. One troubling trend is that both honey bee and *Varroa* proteomics samples consistently result in lower peptide identifications compared to conventional model species, which is hindering research on not only social disease-resistance mechanisms and honey bee-mite interactions, but countless other biological topics.

We begin by conducting a proteogenomics interrogation to suggest improvements for both the *Varroa* and the honey bee genome annotations, and to help alleviate the limitations of proteomics technology. The resulting protein databases and web-based protein atlas will serve as resources for future *Varroa* and honey bee proteomics experiments. Next, we investigate the chemical ecological aspects underpinning hygienic behaviour in honey bees (one form of social immunity against parasites like *Varroa*). We use gas chromatography-mass spectrometry to analyze abundances of volatile and non-volatile odorants in freeze-killed and age-matched healthy brood, as well as *Varroa*-infested and non-infested brood. We identified 10 differentially emitted compounds, 2 of which (β-ocimene and oleic acid) are intriguing candidates as hygienic behaviour-inducers based on their previously known functions in honey bees and other social insects. Next, we investigate these two compounds' abilities to induce hygienic behaviour using a series of behavioural assays. We found that, depending on the context, both odorants can

induce hygienic behavior, and they may be acting synergistically. Finally, we begin to investigate physical and biochemical interactions between these odorants and two odorant binding proteins – OBP16 and OBP18 – which are thought to aid in disease odorant detection. We find that β-ocimene is a ligand of OBP16 and oleic acid is a ligand of both OBPs. We conclude by beginning to develop RNAi and transgenic methods for investigating the roles of these proteins *in vivo.* Overall, these studies are starting to reveal the simple molecular mechanisms underlying a complex social immunity trait in honey bees.

Lay Summary

Honey bees are important pollinators for agriculture, and the parasitic *Varroa* mite is the most pervasive and economically devastating pest they face. However, our knowledge of *Varroa* biology and how honey bees can resist *Varroa* is limited. Here, we develop community-wide resources for investigating honey bee and *Varroa* biology in the 'omics era. We then use some of these tools to support our own research into one honey bee disease-resistance characteristic: hygienic behaviour. We identify odorant molecules that stimulate the behaviour and suggest a mechanism of action involving interactions with molecular receptors in the bees' antennae. Finally, we begin to develop methods for altering the genes for these receptors to study how they may affect the bees' sense of smell. Overall, these results are helping us understand the simple mechanics behind a complex behaviour and further validate existing approaches for selectively breeding hygienic stock.

Preface

I wrote the entirety of this thesis with the exception of short sections of the Introduction, which I co-wrote with Dr. Judith Trapp. I designed or helped design all the experiments, with the exception of the controlled, cross-species proteomics comparison (Figure 2.6 B and C), which was designed and executed by Sarah Michaud. Biological samples for the same experiment were supplied by Jenny Moon, Dr. George Chung, Dr. Carol Pollock, Dr. Patrick Chung, and Dr. Nicholas Scott, while I analyzed and interpreted the results. For all proteomics experiments in this thesis, Dr. Nikolay Stoynov maintained the mass spectrometry instruments.

I collected all *Varroa* samples, prepared the protein extracts for mass spectrometry, analyzed the data, and interpreted the data. Dr. Queenie Chan developed and coded the architecture for the *Varroa destructor* Protein Atlas, which was built using the data I produced, analyzed, and interpreted.

I collected all honey bee brood samples for gas chromatography-mass spectrometry. I performed the antenna proteomics experiments and analyzed and interpreted the data. I helped design the electroantennography tests with Troy Collins and Ylonna Gallagher-Kurtzke, who executed them. Lina Madilao maintained the gas chromatography-mass spectrometry instruments. I designed and performed all behavioural assays with assistance from Heather Higo, Abigail Chapman, and Leonard Foster. I helped design the *in vitro* binding assays with Dr. Immacolata Iovinella and Dr. Paolo Pelosi, who conducted the assays. I performed all RNAi and transgenics

experiments, with assistance from Abigail Chapman. Dr. Martin Beye, Dr. Marianne Otte, and Dr. Christina Schulte were instrumental in training me on the transgenic techniques.

Research on honey bees and *Varroa*, which are non-cephalopod invertebrates, does not require animal ethics approval at the University of British Columbia.

Excerpts from Chapter 1 have been published.

McAfee, A. and Foster, L. J. (2016). Proteogenomics: Recycling public data sets to improve genome annotations. *Methods Enzymol*. 585:217-43. I wrote the manuscript and made the figures, with assistance from L. J. F.

Trapp, J.*, McAfee, A.*, and Foster, L. J. (2016). Genomics, transcriptomics and proteomics: enabling insights into social evolution and disease challenges for managed and wild bees. *Mol. Ecol.* 26:718-39. *Authors contributed equally. I made the figures and co-wrote the manuscript with J. T., with assistance from L. J. F.

A version of Chapter 2 has been published.

McAfee, A. Michaud, S., and Foster, L. J. (2015). A controlled, cross-species dataset for exploring biases in genome annotation and modification profiles. *Data Brief.* 5:829-33. I wrote the manuscript and S. M. produced the dataset. Grants to L. J. F. funded the research. L. J. F. edited the manuscript.

McAfee, A. Harpur, B. A., Michaud, S., Beavis, R. C., Kent, C. F., Zayed, A., and Foster, L. J. (2016). Towards an upgraded honey bee (*Apis mellifera* L.) genome annotation using proteogenomics. *J. Proteome Res*. 15:411–21. I wrote the manuscript while B. A. H. and

C. F. K. analyzed RNA sequencing data, S. M. produced some of the mass spectrometry data, B. R. C. cataloged non-tryptic cleavages, A. Z. provided the genome-sequenced bee. Grants to L. J. F. funded the research. L. J. F. helped design experiments and edited the manuscript.

McAfee, A., Chan, Q. W. T., Evans, J., and Foster, L. J. (2017). A *Varroa destructor* protein atlas reveals molecular underpinnings of developmental transitions and sexual differentiation. *Mol. Cell. Proteomics*. 16:2125-37. I wrote the manuscript and produced the figures. Q. W. T. constructed the Protein Atlas and J. E. provided the draft genome annotation. Grants to L. J. F. funded the research. L. J. F. helped design experiments and edited the manuscript.

A version of Chapter 3 has been published.

McAfee, A., Collins, T. F., Madilao, L. L., and Foster, L. J. (2017) Odorant cues linked to social immunity induce lateralized antenna stimulation in honey bees (*Apis mellifera* L.). *Sci. Rep*. 7:46171. I wrote the manuscript and produced the figures. T. F. C. conducted the electroantennography experiments and L. L. M. maintained the gas chromatography mass spectrometers. Grants to L. J. F. funded the research. L. J. F. helped design experiments and edited the manuscript.

A version of Chapter 4 and excerpts from Chapter 5 have been published.

McAfee, A., Chapman, A., Iovinella, I., Gallagher-Kurtzke, Y., Collins, T. F., Higo, H., Madilao, L. L., Pelosi, P., and Foster, L. J. (2018). A death pheromone, oleic acid, triggers hygienic behaviour in honey bees (*Apis mellifera* L.). *Sci. Rep.* 8:5719. I wrote the

manuscript and produced the figures, with the exception of Figure 5.1, which was first produced by Dr. Immacolata Iovinella, then was edited by myself. A. C. and L. J. F. assisted with behavioural assays. Y. G. K. conducted the electroantennography experiments. T. F. C. established the initial electroantennography methodology. H. H. helped design the behavioural assays. I. I. and P. P. performed the ligand binding assays. L. J. F. provided guidance, feedback and funding.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BCA	Bicinchoninic acid
BEP	Brood ester pheromone
BLAST	Basic local alignment search tool
dsRNA	Double-stranded RNA
DSHB	Developmental studies hybridoma bank
EAD	Electroantennographic detection
EAG	Electroantennography
ESI	Electrospray ionization
FDR	False discovery rate
FKB	Freeze-killed brood
GFP	Green fluorescent protein
GC-MS	Gas chromatography-mass spectrometry
GO	Gene ontology
GSR	Gene score resampling
HPLC	High pressure liquid chromatography
HSD	Honestly significant difference
HSP	Heat-shock protein
IC50	Inhibitory concentration (50%)
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IQR	Interquartile range
L5	Fifth instar larval stage

LC	Liquid chromatography
LFQ	Label-free quantification
MS	Mass spectrometry
MSMS	Tandem mass spectrometry
NCBI	National center for biotechnology information
NPN	N-phenyl-1-naphthlamine
OBP	Odorant binding protein
OGS	Official gene set
ORF	Open reading frame
PCR	Polymerase chain reaction
PTM	Post-translational modification
QFF	Q fast flow
Q-TOF	Quadrupole-time-of-flight
RNAi	RNA interference
RT-PCR	Reverse transcription polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPME	Solid phase microextraction
STAGE	Stop and go extraction
TCEP	tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
VSH	*Varroa*-sensitive hygiene

Glossary

Hygienicity	The degree of hygienic behaviour aptitude, as determined by conducting a freeze-killed brood assay. Usually on a scale of 0 – 100%
Semiochemical	A pheromone or other chemical that conveys a signal from one organism to another so as to modify the behaviour of the recipient organism
T2A peptide	A self-cleaving peptide sequence originating from the Thosea asigna virus 2A

Acknowledgements

I will be forever indebted to my supervisor, Leonard Foster, who has provided me with intellectual, emotional, and financial support since I first joined the group as an undergraduate. Among other things, he encouraged my pursuit of freelance writing, taught me how to achieve a sustainable work-life balance and has been an inspiring champion for combatting gender bias at the highest level. It has been a pleasure to conduct my training in his laboratory.

My committee members – Dr. Erika Plettner, Dr. Chris Overall, and Dr. Michael Gordon – have always provided invaluable feedback on my research and helped guide my thesis over the years. Special thanks go to Heather Higo, for mentorship and inspiration, and Abigail Chapman, for her assistance over the last two summers. Thank you also to all those who have helped complete various aspects of this work, including Jenny Moon, Dr. Queenie Chan, Ylonna Gallagher-Kurtzke, Sarah Michaud, Troy Collins, Dr. Nikolay Stoynov, Dr. Immacolata Iovinella, Dr. Christina Schulte, Dr. Marianne Otte, Dr. Brock Harpur, Dr. Amro Zayed, Derek Micholson, Dr. Jay Evans, Dr. Marla Spivak, and Dr. Danica Baines, Dr. Sam Chen, Dr. Chris Tam, Dr. Craig Kerr, Dr. Anthony Khong, Sally Greenwood, and many, many more along the way.

Thank you, Veronique Le Blanc, Madeline Couse, Michelle Kwon, Craig Kerr, and Jason Rogalski or your enduring friendship, support, and tough love.

I have eternal gratitude to my loving parents, Leonard McAfee and Debra Richards, who have given me the freedom to set my own trajectory and the support to keep it steady.

And of course, I could not have completed my thesis without the help from many sources of funding, including the Natural Sciences and Engineering Research Council, the University of British Columbia, the Michael Smith Foreign Studies Supplement, the Australasian Proteomics Society, the Canadian Association of Professional Apiculturists, the Avtar S. Atwal scholarship for Canadian pollinator research, and the BC Proteomics Network. Thank you for providing me the financial support to create and disseminate this body of work.

Dedication

To Dr. Margaret Nishikawara.

The late Margaret Nishikawara was my great aunt, who passed away on June 8th, 2014. As a Japanese-Canadian woman, she admirably completed her Bachelor of Arts, Master of Arts, and Doctor of Philosophy (in Biochemistry) degrees amidst an era of war-fueled discrimination. Still, she achieved tenure in 1954 at the Ohio State University (OSU) in the Department of Physiology and Cell Biology. She held it for 32 years before retirement. She specialized in enzymology and contributed important work to OSU's Commission on Women and Minorities in the 1970s.

Margaret's prolific career in science and generosity in the face of deep-rooted social challenges has inspired me throughout my degree. Thanks to pioneers like her, there are brighter futures for women in science.

Chapter 1: Introduction[1]

1.1 Honey bees' role in agriculture and the environment

Humans have been intrigued with honey bees since the time of our early ancestors. Spanish cave paintings from 8,000 years ago depict humans collecting honey from wild bee hives, and *A. mellifera* (now known as the Western honey bee) was domesticated as early as 2600 BC[1]. Honey bees are native to Africa, Europe, and parts of Asia, but since domestication, they are now found on every habitable continent (Figure 1.1)[2]. They provide us with a host of natural products – honey, wax, and propolis, to name a few – but most importantly, today they are one of the primary means of crop pollination[3]. In the United States alone, honey bee pollination services are estimated to be worth between $1.6 and $14.6 billion annually[4,5] and in Canada, estimates are between $3 and $5.5 billion[6]. Indeed, colonies are employed in 90% of agricultural operations requiring actively managed pollination, representing about 35% of crops[7]. Honey bees are clearly essential for the agricultural sector; however, in their endemic range, they also provide keystone pollination services for the natural ecosystem, along with the approximately 20,000 other bee species and countless other insect and non-insect pollinators[8]. Furthermore, within the realm of basic science, honey bee colonies are a prime model system for studying the genetics and evolution of complex social behaviour[9,10]. Probably the best-known example of such a behaviour is the waggle dance language, the decoding of which won Karl von Frisch a Nobel Prize in 1973[11,12].

[1] Content in this chapter has been published: McAfee, A. and Foster, L. J. Proteogenomics: Recycling public data sets to improve genome annotations. *Methods Enzymol.* 585:217-43. Copyright (2016) Elsevier. Reprinted (adapted) with permission; Trapp, J.*, McAfee, A.*, and Foster, L. J. Genomics, transcriptomics and proteomics: enabling insights into social evolution and disease challenges for managed and wild bees. *Mol. Ecol.* 26:718-39. Copyright (2016) John Wiley & Sons. Reprinted (adapted) with permission. *Authors contributed equally.

Within the last decade, North American beekeepers have experienced unusually high overwintering colony losses of 25% annually (averaged over the last 11 years)[13] – up to 50% in some Canadian provinces[14]. These losses are caused by several interacting factors, including impacts of agrochemical exposure, harsh winters, diseases, parasites, and other challenges[8,14]. (Figure 1.2). Beekeepers work to overcome these challenges through nutritional supplementation, splitting colonies in the spring, diligent pest and pathogen monitoring, and medication with antibiotics, acaricides, and fungicides when needed. Honey bee populations have undergone regional declines, mainly within the United States and Europe[8,15], with politics and socioeconomic fluxes (which heavily influence the number of beekeepers in a given region) being the main drivers[16]. However, on a global scale, the honey bee population is steadily increasing – since 1961, global honey bee colony numbers have increased by about 45%[8]. While it is encouraging, this rate of population growth is far below the increase in agricultural pollination demand for the same period (>300%)[8]. What's more, the colony growth that beekeepers have achieved is at a great economic cost (*e.g.* the cost of acaricides, queen replacement, labour, *etc.*) which many beekeepers consider to be unsustainable. Therefore, research supporting colony health and disease resistance, including improved basic knowledge of the biological mechanisms underlying this resistance, are in demand.

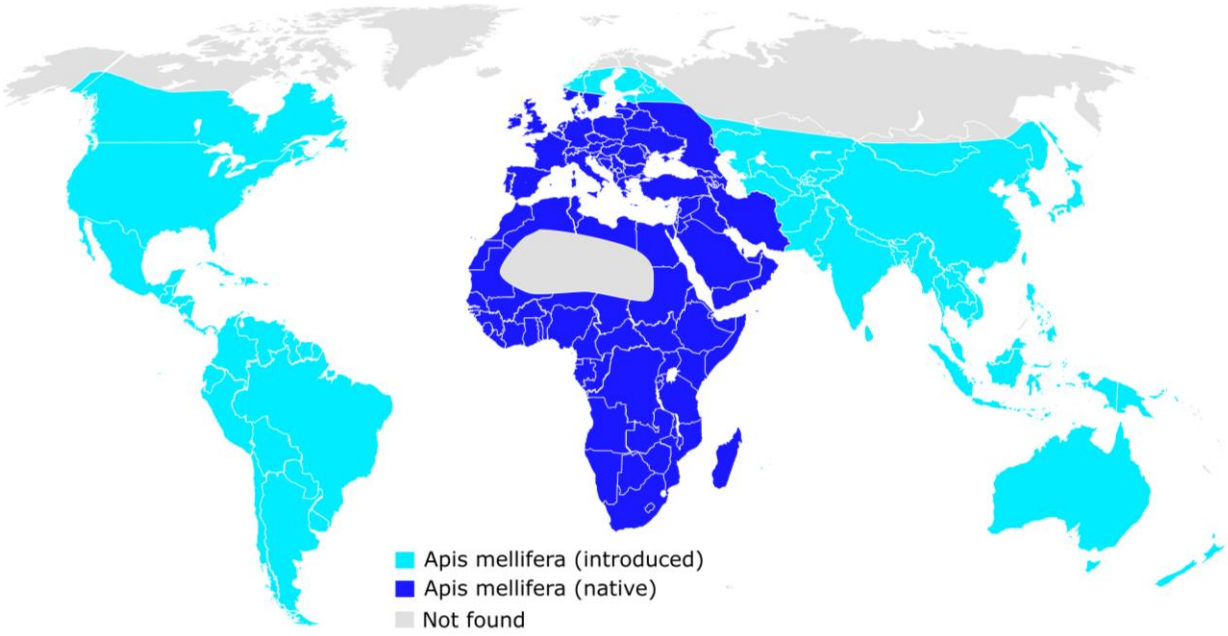


Figure 1.1. Honey bee global distribution

The geographic distribution of honey bees has expanded greatly as a result of human activities. Honey bees were once found only in Europe, the Middle East and Africa (except the Sahara Desert), but now are found on all continents except Antarctica.

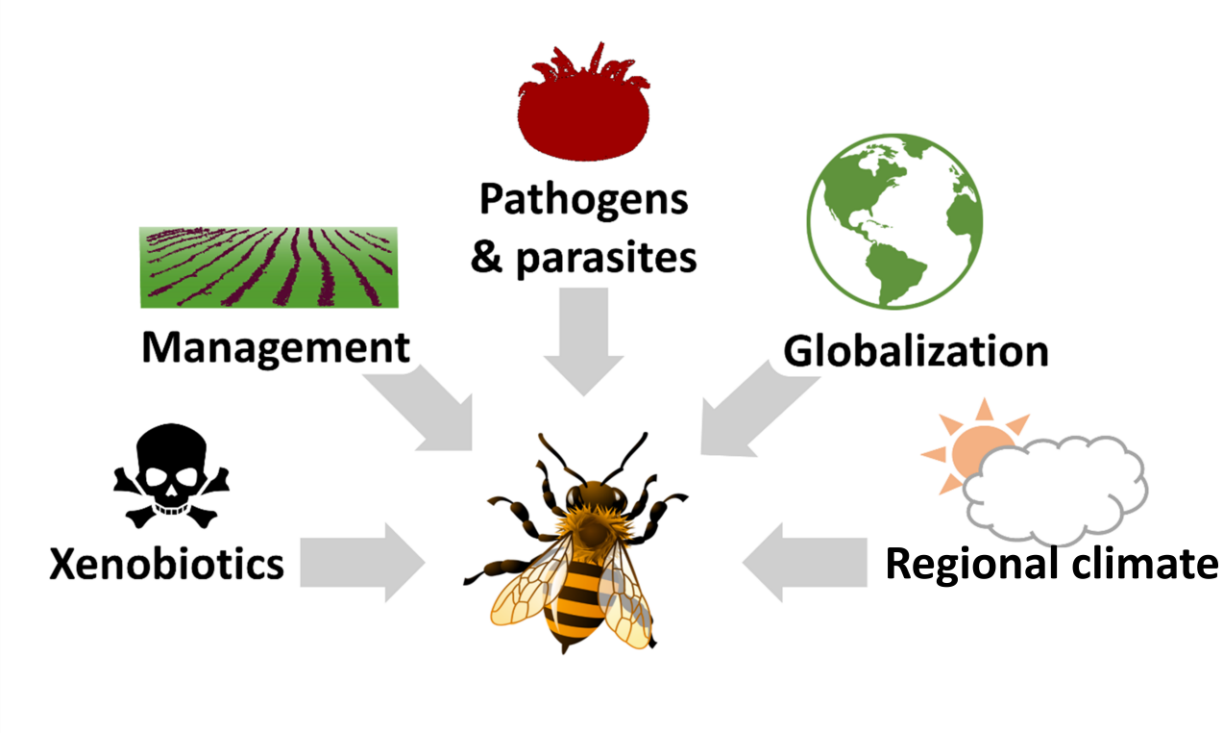


Figure 1.2. Major drivers of honey bee colony health and productivity

Many factors interact to influence colony health. For example, honey bees are commonly exposed to neurotoxic insecticides and other agrochemicals (xenobiotics), which can increase their susceptibility to queen loss and some diseases. Colony management strategies in an industrial setting (*i.e.* high densities of colonies and apiaries in one geographical region) can promote the spread of disease within and between apiaries. A wide variety of pests and pathogens can infect bees, ranging from viruses, bacteria, and fungi to parasitic mites. The spread of some pathogens has occurred inadvertently as a result of globalization (*i.e.* frequent import and export of live honey bees and their products between nations). Other environmental factors, such as the local climate, can impact colony health (*e.g.* harsh Canadian winters).

1.2 Brood pathogens and parasites

Eusocial insects live in teeming societies with thousands of their kin. In this crowded environment, the risk of pathogen and parasite transmission is high. With the exception of *Nosema apis* and *N. ceranae*, most diseases preferentially affect the brood stage (*i.e.* larvae and pupae) or are transmitted by brood-associated parasites. Brood diseases and parasites – in particular *Varroa destructor*[17-20] – and management thereof, are one of the biggest challenges beekeepers face today.

1.2.1 ***Varroa destructor***

The *Varroa destructor* mite (class: Arachnida; hereon referred to as *Varroa*) is the most devastating pest for Western honey bees[17,20,21]. This obligate ectoparasite feeds on honey bee hemolymph (blood), simultaneously weakening its host, suppressing the innate immune system, and transmitting debilitating viruses[22] such as Israeli acute paralysis virus (IAPV), acute bee paralysis virus (ABPV), and deformed-wing virus (DWV)[23-26]. Of these, DWV is by far the most prevalent and recent work has shown that it exacerbates *Varroa*-induced immunosuppression of the honey bee, further benefitting *Varroa* by enhancing mite reproduction[25].

Varroa's natural host is the Eastern honey bee (*A. cerana*), and millions of years of co-evolution has led *A. cerana* to develop tolerance mechanisms to minimize the mite's negative impact on the colony[27-30]. Traits that enable *A. cerana* to resist *Varroa* include shorter brood development, entombment of mite-infested brood, systematic cap perforations, and importantly, a mechanism of social apoptosis[27-30]. Social apoptosis refers

to how *A. cerana* brood is more easily damaged by parasitization, causing developmental delay and even death; therefore, self-sacrifice of an individual benefits the colony by disrupting the *Varroa* reproductive cycle.

Varroa is well-tolerated by *A. cerana*, but by the mid-1900s the mite shifted hosts to include both *A. cerana* and *A. mellifera*. *A. mellifera* is the species that is most commonly used for active crop pollination today and is less effective at defending itself[22,28,30]. *Varroa* was first detected in North America in 1987, and today, it is found ubiquitously with the exception of Australia and some isolated islands[22]. Managed *A. mellifera* colonies infested with *Varroa* have significantly shorter lifespans than uninfested colonies unless they are actively treated with miticides[31,32], causing serious negative economic impacts[3,19,33].

The most common commercial miticides include synthetic compounds (coumaphos, tau-fluvalinate, and amitraz) as well as organic acids (formic acid and oxalic acid). In the United States, *Varroa* gained resistance to tau-fluvalinate and coumaphos as early as 1998 and 2003, respectively[34], and beekeepers are increasingly worried about emerging resistance to amitraz. The organic acids continue to be effective miticides, but their use requires specialized training and personal protective equipment. Furthermore, their mechanism of action is not well understood. The general threat of emerging resistance has fostered growing interest in honey bees that are naturally resistant to mite infestations[35-41].

1.2.2 **Foulbrood disease**

Prior to the 1970s, the main threat to honey bees in North America was the bacterial disease known as American Foulbrood (AFB)[42], caused by *Paenibacillus larvae*. The disease gets its name from the foul-smelling odour emitted from the infected brood. Brood are only susceptible during the first 48 h following eclosion (egg hatching), when the initial infection establishes in the larval gut[43,44]. Infected brood usually die shortly after their cells are capped and liquify into a brown, viscous, foul-smelling material, eventually forming a hardened, spore-containing scale. AFB spores can remain viable for decades[45].

AFB is exceptionally virulent, such that common practice has historically been to burn any beekeeping equipment that may harbor the bacteria or its hardy spores[45]. Sometimes entire colonies are incinerated, if the infection is sufficiently advanced. Minor infections can be medicated with antibiotics such as oxytetracycline or tylosin; however, these do not inactivate the spores and resistant bacterial strains are abundant. Resistance to oxytetracycline emerged in North America in the year 2000[46], although tylosin appears to still be effective. Again, this resistance arms race spurred breeding programs to select for AFB-resistant honey bees, resulting in the Minnesota Hygienic stock, which is still sought out by beekeepers today[47-50].

1.2.3 **Chalkbrood disease**

Chalkbrood is a fungal brood disease caused by *Ascosphaera apis*[51]. It, too, invades the gut of young larvae via ingested spores. From there, its mycelium grows to occupy the gut cavity, causing starvation, and eventually takes over the larva's body to form a mycelium-coated corpse, or "mummy". Infected larvae typically die soon after cell capping.

Chalkbrood disease is not normally fatal for the colony as a whole, but it can weaken the colony and increase susceptibility to secondary infections[52]; therefore, it is still economically damaging. There is no approved treatment for chalkbrood in honey bees. However, the Minnesota Hygienic stock described above is resistant to chalkbrood in addition to AFB[52,53], and other efforts are being made to identify SNP markers for chalkbrood resistant colonies[54].

1.2.4 **Viruses**

By now, twenty-four different viruses are known to infect honey bees, in addition to several newly identified candidates[55], and at least eight are associated with *Varroa*[42,56]. The most common are sacbrood virus (SBV), Kashmir bee virus (KBV), ABPV, IAPV, and DWV. Of these, DWV is by far the most prevalent[57]. The signs and symptoms of infections caused by these viruses vary greatly, from the characteristic shriveled wings of DWV, to the larval browning of SBV, to the hairlessness and twitching of ABPV and IAPV. Not all of these viruses affect the brood directly, but with the *Varroa* mite as one of their common sources of transmission, it is often at the brood stage when the bees become exposed. Some viruses are lethal to brood (SBV) and adults (KBV, ABPV, and IAPV), while others normally only decrease fitness (DWV), but all can collectively weaken the colony and virus-induced immunosuppression makes the bees susceptible to other diseases[25].

Like chalkbrood, there are no commercial treatments for honey bee viruses. Hunter *et al.*[58] developed a promising RNAi-based, therapy for virus-infected colonies (called “Remebee”);

however, it has not been brought to market. At present, the best way to control the viral load is indirectly through controlling their vector (*Varroa*).

1.3 Immune defenses

1.3.1 Innate immunity

Honey bees lack a T- and B-cell-mediated adaptive immune system, so to combat the strong disease pressure in their colonies, they instead rely on a combination of innate immunity and social immunity traits[59,60]. Innate immunity mechanisms include molecular strategies such as antimicrobial proteins[61], siRNAs[62], and the melanization cascade catalyzed by prophenoloxidase[63]. The RNAi, Jak/STAT, Toll, JNK, and Imd/Relish pathways, which govern these immune responses, are well-conserved in honey bees as with other arthropod and non-arthropod taxa, with well-defined molecular mechanisms (reviewed in Evans *et al.*[59] and Brutscher and Flenniken[62]). Collectively, these mechanisms confer some resistance to bacteria (*e.g.* AFB), viruses (*e.g.* DWV and IAPV), and possibly some fungi (*e.g. Nosema spp.* and chalkbrood); however, none are effective defenses against *Varroa.* Hygienic behavior, a form of social immunity, is the central topic to this thesis and is the focal disease-resistance mechanism from here on.

1.3.2 Social immunity

Dense insect societies have unique challenges for disease management. Ants[64-68], termites[69-71], and honey bees[48,49,72-74] have all evolved social mechanisms of disease resistance which mitigate transmission and reduce the pathogen load (Figure 1.3). The term ‘social immunity’ refers to a collection of social behaviours related to hygiene that maintain colony health[59,60,75-77]. For

example, ants transport dead nestmates to their midden heaps, termites bury or entomb their dead in graves, and honey bees remove dead and diseased brood from the hive. E. O. Wilson described these processes as 'necrophoresis'[64], or the movement of dead individuals away from the colony. Necrophoresis effectively reduces pathogen reservoirs, inhibiting the spread of diseases and parasites from fallen nestmates to those who endure[64,75,78].

Honey bees have many social immunity traits, including (but not limited to) grooming[60], self-medicating with antimicrobial substances[79-81], social fever[82], hygienic behaviour[73,83], and *Varroa*-sensitive hygiene[84,85]. Here, we will focus on the latter two, since these are the main social immunity behaviours against the brood diseases and have historically received the most attention for breeding programs. Both traits are highly desirable because they are heritable and improve colony disease resistance while avoiding costly, laborious, and sometimes dangerous colony medication strategies.

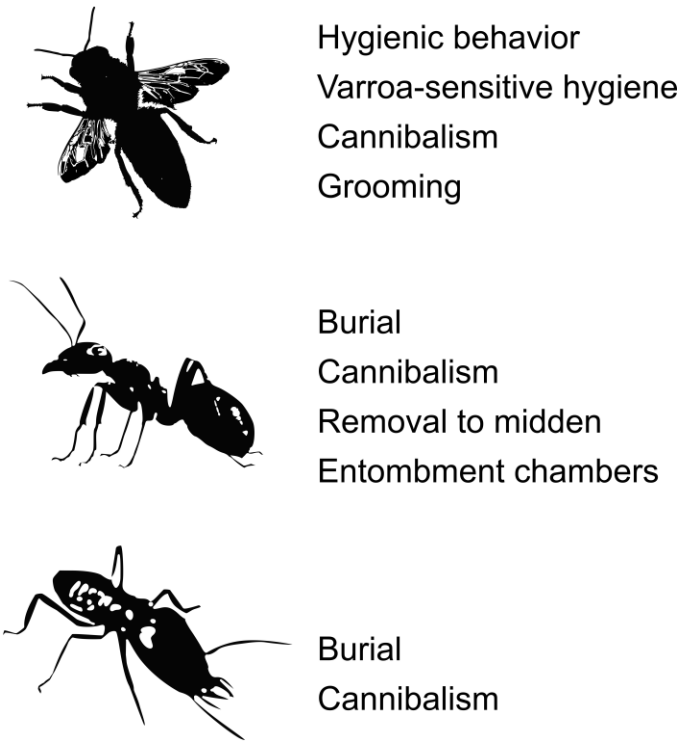


Figure 1.3. Examples of social immunity strategies employed by ants, termites, and honey bees.

1.3.2.1 Hygienic behaviour

Hygienic behaviour is highly heritable[50,86] and holds great economic potential[48,49], but the underlying molecular mechanism has only been partially deciphered. The behaviour itself is characterized by worker bees detecting, uncapping and removing diseased or infested brood (Figure 1.4) and was first described by WC Rothenbuhler in the 1960s[73,83]. Now, we know it is an effective defense against American foulbrood[50], chalkbrood[52], and (in some cases) *Varroa*[87,88], although it is generally accepted that *Varroa*-sensitive hygiene is more effective against the latter[85,89]. Typically, hygienic behaviour is an age-specific task performed by worker bees between 15 and 20 days after emerging as adults (*i.e.* post-nursing and pre-foraging)[90]. Workers can be 'uncappers' (who identify the diseased individuals and chew away the cap) and

'removers' (who may widen the cap further, pull the pupa out of the cell, and eventually remove it from the hive)[91].

Like other social insects, honey bees identify their diseased nestmates via chemical cues[39,91-96]; however, since partway through development (late 5th instar larvae and older) the brood becomes capped and completes development in the confines of a sealed wax cell, the workers have an added challenge. The physical barrier between the bees who execute the behaviour and the brood interferes with their ability to detect their targets. Detecting the dead, diseased, or parasitized capped brood is therefore thought to rely on volatile odorant signals that permeate the wax cell cap[93], but very few hygienic behaviour-inducing odorants have been identified and confirmed behaviourally[94,95]. Swanson *et al.*[94] found that a volatile chalkbrood odorant (phenethyl acetate) induced hygienic behaviour and Nazzi *et al.* showed that one volatile *Varroa*-associated odorant ((Z)-6-pentadecene) does the same[95]. Non-volatile cues have not yet been investigated behaviourally in honey bees, despite including some of the most taxonomically conserved necrophoretic and necrophobic compounds (*e.g.* oleic acid and linoleic acid)[66,69,70,78,97-102].

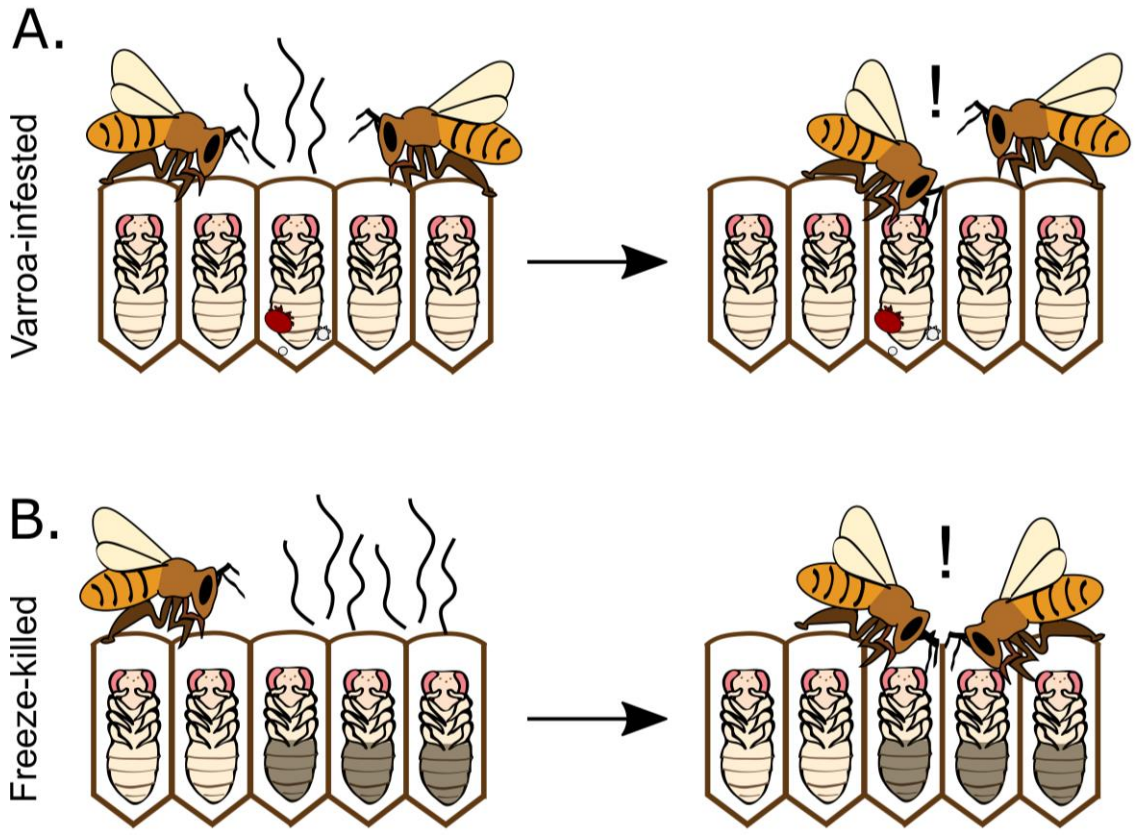


Figure 1.4. Schematic of honey bees executing hygienic behaviour.

A. Hygienic behaviour against *Varroa*-infested brood. Honey bees may uncap and remove infested pupae, or uncap and recap the cell, disrupting the *Varroa* life cycle. B. Hygienic behaviour against freeze-killed brood. The freeze-killed brood assay is the most sensitive method of testing colonies for hygienic behavior.

1.3.2.2 *Varroa*-sensitive hygiene

Varroa-sensitive hygiene, or VSH, is similar to hygienic behaviour, but high VSH colonies are better at targeting *Varroa*-infested brood, specifically. VSH is related to the suppression of mite reproduction (SMR) trait, but may or may not include a brood effect (*i.e.* a brood-mediated mechanism for suppressing mite reproduction, rather than adult-mediated, recently termed "social apoptosis")[85,89], depending on one's definition. Like hygienic behaviour, VSH bees are thought to be able to detect mite-infested brood via volatile odorant cues[95], which could be a

result of infestation-induced developmental delay of the brood[96], stronger odorant signals emitted from self-destructing brood[30], superior olfactory sensitivity of the workers[39], or some combination of these factors.

1.3.2.3 Selective breeding for social immunity

Like other agricultural livestock, honey bees have long undergone selective breeding for beneficial traits (*e.g.* gentleness). While Rothenbuhler[73,83] was the first to systematically breed hygienic honey bees, Spivak and Downey were the first to mechanize how to do it on a large scale[47]. They helped develop and validate a robust field test called the "freeze-killed brood (FKB) assay" which is still used for identifying and selecting hygienic colonies to this day. The method has improved in sensitivity and efficacy over time – now, it involves freeze-killing patches of brood with liquid nitrogen, returning them to the colony, then scoring each colony based on the percent removal (and partial removal) of the dead brood[48]. Ideally, the test is repeated at least twice on different dates and the average percent removal across all tests yields the hygienic score. Highly hygienic colonies are generally considered to be those that score ≥95% on this test, and colonies selected in this way are more resistant to AFB, chalkbrood, and sometimes *Varroa*[48,49,88]. Spivak and others used this assay to produce the Minnesota Hygienic honey bee stock which were once distributed across North America; however, the breeding program is no longer active[103].

VSH is a more difficult trait to select for, since it requires experimentally adding mites to individual brood cells, or estimating how the mite infestation level changes in a donor frame after incubating in the test colony[84]. Nevertheless, some research groups have succeeded in

producing high VSH stock[104,105]; however, the field-assay selection methods are generally not appropriate for use in an industrial setting due to their tedious nature.

Unfortunately, the FKB method of selecting for hygienic colonies has also not been widely adopted by the honey bee breeding industry, owing mainly to the time and labour involved. Therefore, marker-assisted selection may be a more effective breeding tool[37,105,106]. Some research has identified genetic markers for social immunity[107-109]; however, since honey bees have extremely high rates of genetic recombination[110], non-causal genetic markers quickly lose their linkage with their causal partners. Therefore, we expect that expression biomarkers (*i.e.* proteins or transcripts) will be more durable than genetic markers.

Differential expression studies are the typical starting point for identifying protein and transcript biomarkers. Differentially expressed proteins in the larval cuticle of high and low VSH colonies suggest that biochemical events in the larvae which are later targeted for removal may help stimulate the behaviour[38], supporting the notion of the brood effect that had previously been observed[85,89]. In addition, multiple gene products involved in chemoreception are up-regulated in antennae from highly hygienic honey bees[38,39,106,111-113]. For example, members of the take-out/juvenile hormone binding protein superfamily, which are soluble receptors in the hemolymph[38], and several odorant-binding proteins (OBPs) were upregulated in hygienic Western[39,106,112,113] and Eastern[114] honey bees, including OBP3, 4, 13, 14, 16, and 18. OBPs carry odorants to the sensilla nerves and stimulate the odorant receptors; therefore, they likely enhance the detection of odorant molecules associated with disease or dead bees[106].

Previously, our own research group developed a panel of protein biomarkers for selectively breeding hygienic honey bees[112]. This was achieved by correlating antennal protein expression with the hygienic score and combining the 9 best biomarkers into a predictive model. Since then, these markers have been validated as effective breeding tools, yielding comparable results in terms of field assay scores and disease challenge outcomes as the conventional selection method[112]. OBP16 and OBP18 are two of these biomarkers[106]. While most previous differential expression studies analyze only a handful of samples in each experimental group, the correlational analysis that produced these 9 biomarkers included 167 colonies across two geographically distinct sites. Therefore, of all the OBPs that have been implicated in hygienic behaviour in the past, these two have the strongest supporting evidence, and in the case of OBP18, corroboration in an independent study[115].

1.4 Honey bee olfaction

The antennae are the sites of peripheral odorant detection in the honey bee[116]. They are covered with small shield- and hair-like structures (sensilla), each of which houses the dendrites of 5 to 35 olfactory receptor neurons (ORNs)[117,118]. Bundles of ORN axons project to the antennal lobe at the base of the antenna, where nerve signals are transmitted to other parts of the brain, such as the lateral horn and the mushroom body[119]. The antennal lobe itself is made up of 165 glomeruli, which is roughly the same number of predicted olfactory receptors (ORs), suggesting that each ORN expresses one OR and is joined to one glomerulus[117]. The ORs are located on the ORN membrane and are in contact with the sensillium lymph fluid that occupies the volume between the cuticular wall and the ORNs.

1.4.1 **Role of odorant binding proteins**

It is through the sensillium lymph fluid that OBPs are thought to aid in the transport of volatile odorants from the external olfactory pores to the ORNs[120-123]. They could also bind and sequester abundant odorant molecules, making them unavailable to stimulate the olfactory nerves and therefore allowing less abundant odorants to be sensed[121,124]. OBPs are often expressed in non-olfactory tissue, where they may act as general carriers for small molecules for other biological processes[120], or possibly play a role in the secretion and release of pheromones[122]. Very little is known about the specific roles that different honey bee OBPs have in odour discrimination and sensitivity. Furthermore, although expression of OBP16 and OBP18 are strongly correlated with hygienic behaviour, the functional relationship between these OBPs and performance of the behaviour has not been previously demonstrated.

Like other OBPs, OBP16 and OBP18 are soluble, short proteins 135 amino acids long. They are in the C-minus OBP sub-family, meaning that they lack the 2nd and 5th cysteine residues involved in conserved disulfide bonds[120]. They are expressed most strongly in the antennae and legs, as well as somewhat in parts of the cuticle (tergite and sternite)[120,125]. However, the legs and cuticle of the honey bee are not known to be involved in olfaction and do not appear to express ORs[117]. Binding and release of odorants from OBPs is thought to occur in a pH-dependent manner, with binding generally taking place at the low pH, and release taking place at higher pH (*e.g.*, near the ORs)[126]. The interaction between the ligand and the OR is what stimulates the action potential in ORNs – a signal which is transmitted to the brain structures for higher-order processing, eventually leading to the perception of smell.

1.5 Honey bee research in the 'omics era

Since the initial reports of high over-wintering colony losses, there has been increasing interest in probing relevant aspects of honey bee biology such as immunity and interactions with pathogens[127,128] or pesticides[129-133] to identify the root causes. Because honey bees are a relatively recent subject of molecular biological research, there are few tools available (*e.g.*, there are very limited cell line resources, antibodies, and genome editing techniques) for probing molecular functions compared to model systems. The initial sequencing and annotation of the honey bee genome in 2006[10] and the genome upgrade in 2014[9] have facilitated many of these efforts by making modern biochemical techniques such as RNA-seq, microarrays and mass spectrometry-based proteomics more feasible for researchers.

1.5.1 Genome sequencing and annotation

The honey bee genome was among the first of the arthropod genomes to be sequenced, after *Drosophila melanogaster*[134], *Anopheles gambiae*[135], and *Bombyx Mori*[136] (commonly known as the fruit fly, mosquito, and silk worm, respectively); however, sequencing and assembling it was a surprisingly challenging task. Long AT-rich regions made it difficult to achieve high genome coverage during the first sequencing efforts[9,10]; later, as massively parallel sequencing methods arose, additional genome data were produced using a combination of long read (Roche 454) and short read (ABI SOLiD) technologies to produce a more contiguous genome assembly[9]. Having a complete genome is crucial for understanding honey bee biology, but this knowledge alone cannot resolve all its complexity. First, the genome must be structurally and functionally annotated. Gene structures, such as termini and splice sites, must be catalogued (structural

annotation), and gene functions, such as enzymatic activities and binding partners, must be defined (functional annotation).

The unusual genetic properties of the honey bee genome also complicated the genome annotation procedure **(**Figure 1.5**)**. Gene structure prediction algorithms have inherent limitations[137] – they are usually tuned optimally for the specific genomic properties of the classical model organisms, such as gene size, codon biases, and consensus splice sites[138,139]. If a genome deviates substantially from the usual properties, and if the genes it contains have low sequence similarity to those in previously annotated species, the algorithms drop in sensitivity and specificity[137,138]. The GC content in the honey bee genome is very low (33%) compared to other well-studied species (*e.g.* 41% in *Homo sapiens*, 42% in *D. melanogaster* and 45% in *A. gambiae*; source: NCBI genomes), with some regions as low as 11%. This is reflected by the fact that honey bees also have differential codon usage compared to other arthropods[140]. At least in part because of this, traditional annotation pipelines were insufficient, resulting in the first honey bee gene set containing fewer genes than expected (OGSv1.0; 10,157 genes). *D. melanogaster*, *e.g.*, has a 150 Mb genome, compared to 236 Mb in honey bees, but has 14,692 annotated protein-coding genes[141]. This spurred a second in-depth annotation to create OGSv3.2 – a version with 15,314 protein-coding genes but only 4,083 sequences fully retained from OGSv1.0 – when the updated genome became available. Additional genes in OGSv3.2 are the result of both the improved genome build and an annotation process that is based heavily on biological data that had been generated since the first sequencing (primarily transcript evidence[9]), rather than mainly gene orthologs. However, while many genes were gained in this annotation, many were also discarded without clear evidence that they were incorrect.

Importantly, the quality of the gene annotation, and therefore the protein sequence database, has a profound influence on the utility of mass spectrometry-based proteomics. Since most proteomics approaches rely on having an accurate set of potential protein sequences against which peptide mass spectra can be matched, if the database is missing sequences that exist biologically, they will never be detected in a typical shot-gun proteomics experiment and will not contribute to the biological interpretation and insight gained from that sample.

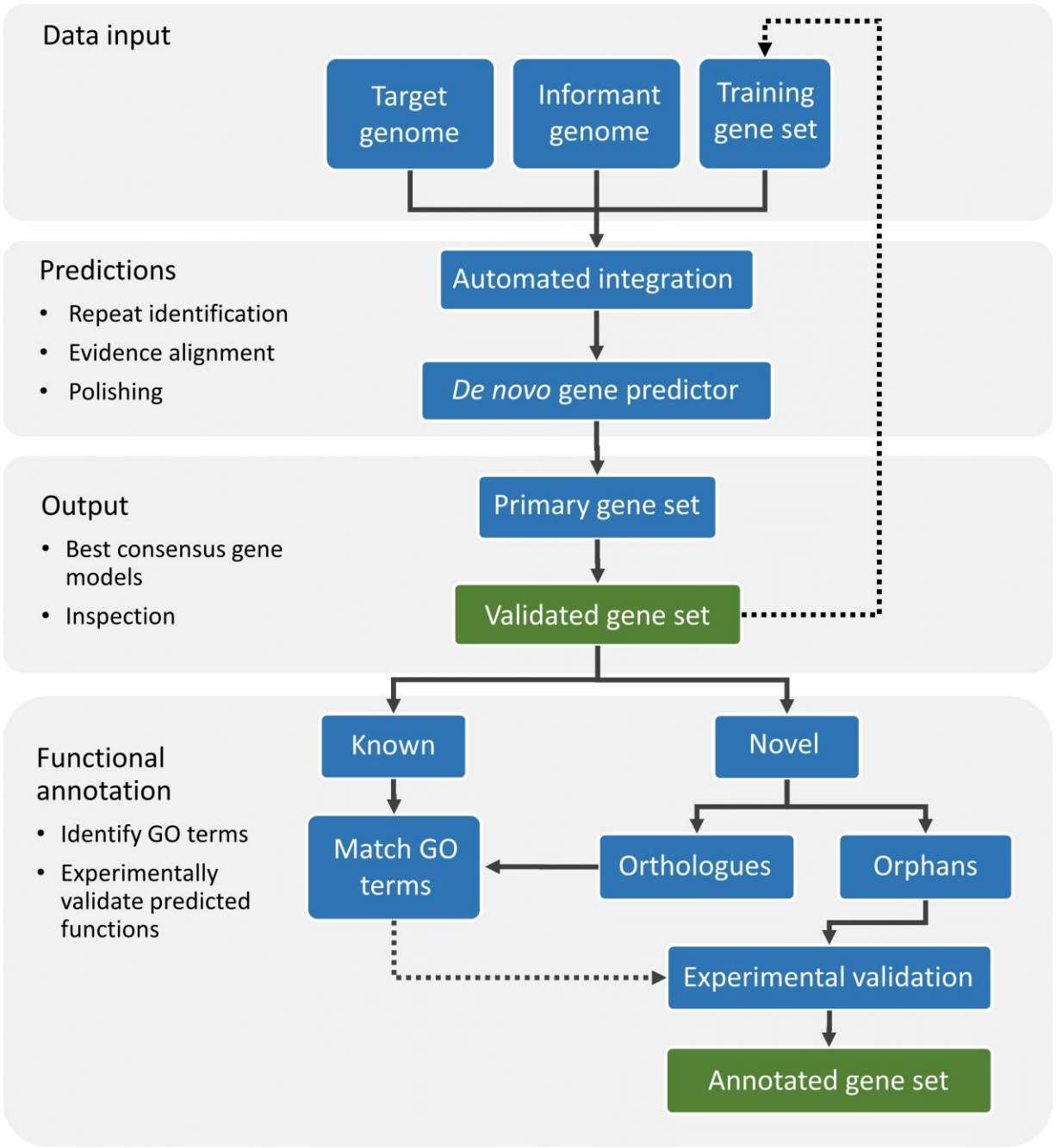


Figure 1.5. Roadmap of a typical gene annotation and functional assignment strategy.

Dotted lines represent optional steps. For example, the validated gene set is an ideal input training gene set for future annotation efforts (*e.g.* using an upgraded target genome sequence), if needed. Additionally, GO terms linked to orthologues of new genes are often directly transferrable, although ideally they should be experimentally validated because some orthologues have been previously shown to perform different functions in the respective species[142].

1.5.2 **Advances in honey bee proteomics**

The field of honey bee proteomics has advanced very quickly and has been repeatedly used to tackle questions rooted in social evolution and disease. Some of the earliest shot-gun proteomics was done as soon as the first bee genome was released, with Chan *et al.*[143] investigating differences in hemolymph (blood) composition between honey bee castes and Hummon *et al.* performing neuropeptidomics[144]. Even in low-throughput proteomics, experiments were still fruitful; for example, comparing brain[145], fat body[146], or whole bee[147] proteomes of forager and nurse workers showed that foragers had reversibly elevated levels of enzymes related to glycolysis and lipid metabolism, presumably due to high energy demands of foraging flights. These differences were found despite only identifying 47, 147 and 81 total proteins, respectively, and support earlier microarray work on the subject[148]. But proteomic technology has advanced impressively during the past decade; *e.g.*, when the proteomes of worker brains were again analyzed[149] spanning the nurse and forager ages, 10-fold more proteins were identified than before. Then, Chan *et al.*[125] published a landmark paper which quantitatively analyzed 29 isolated tissues crossing all three castes, identifying 2,288 proteins. Most recently, Hu *et al.*[150] claimed to achieve honey bee proteome coverage of more than 55% (8,609 proteins) across three tissues (hemolymph, mushroom body and antenna) – the deepest coverage in a single bee data set to date – however, they clearly failed to follow the conventional rules of parsimony in their analysis. Therefore, these numbers are a vast over-estimate. The data is publicly available and re-searching it with the standard parsimonious proteomics workflow and a similar protein database yields just over 3,000 proteins. These numbers still represent some of the highest coverage that has been achieved in the field, which is far lower than what we should expect given the number of annotated proteins and the capabilities of modern technology[151].

Proteomics has now become an indispensable tool for identifying biomarkers for marker-assisted selection. Unlike honey bee SNPs, protein markers are expected to be stably associated with traits through many generations and are less prone to degradation than transcripts. However, our inability to achieve deep proteome coverage is worrisome, as it suggests that there could be many strong biomarkers that are invisible to our analyses. Indeed, of the 13 biomarkers that are currently employed for honey bee selective breeding[112], one was actually a protein that was present in the earliest gene annotation (OBSv1.0), but was heavily edited in the upgraded annotation (OGSv3.6). This protein would have been missed as a biomarker altogether if solely the upgraded annotation database had been used. For these reasons, improving the current annotation will be essential in order to maximize the knowledge gained from past, present, and future honey bee proteomics studies.

1.5.3 **Proteogenomics for improving genome annotations**

Proteogenomics – the process of using peptide or protein information to help inform structural gene annotations – is a data-driven annotation scheme that utilizes high-throughput data from every level of the Central Dogma (Figure 1.6) to confirm or edit existing gene models as part of an iterative refinement process. An example of proteogenomic refinement of the already highly scrutinized human gene set comes from Wilhelm *et al.*[152], who identified 193 novel protein-coding regions, 210 novel protein termini and 40 cases of exon extensions using mass spectrometry data. Furthermore, Kim *et al.*[153] also identified peptides matching to 140 regions formerly annotated as pseudogenes and 9 as non-coding (nc)RNAs. Proteogenomics has also been used to refine the annotations of model organisms, including *Caenorhabditis elegans*[154],

Drosophila melanogaster[155], *Arabidopsis thaliana*[156] and *Mus musculus*[157]. These projects demonstrate that proteomics and transcriptomics data are complementary and they should be used together to create reliable gene models. Since the Kim *et al.* and Wilhem *et al.* publications, stricter guidelines have been asserted for thoroughly annotated species, such as humans, in order to minimize erroneous reporting of novel protein and peptide sequences[158]. However, for newly annotated species like *Varroa* and honey bees, we suggest that a less strict approach can be appropriate. This is particularly relevant if the purpose is to improve the training gene set for future iterations of structural annotations, rather than to define "new genes" outright, directly from the protegenomics data.

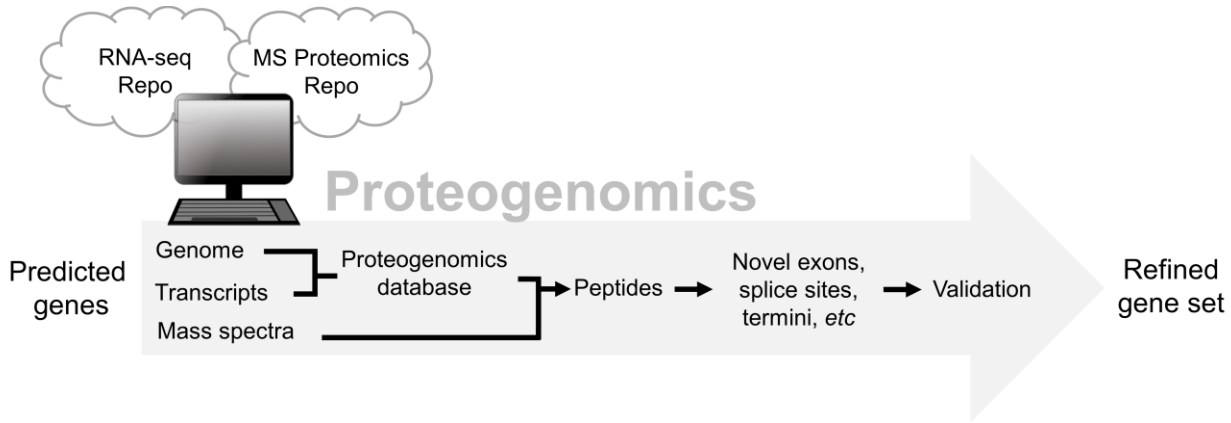


Figure 1.6. Schematic of the proteogenomics workflow.

Input data may be generated in-house or downloaded from public repositories. Mass spectrometry data is searched against the proteogenomics protein database to identify novel peptides, which indicate unannotated gene features.

Genetic elements such as new splice sites, protein termini, exons, alternative reading frames and sometimes entirely new genes[159-161] can be identified by searching mass spectra against a six-frame genome translation, three-frame transcriptome translation, or other custom databases.

Since it offers a relatively unbiased survey of the genomic landscape, this workflow is particularly useful for species with genomic properties that are ill-suited for conventional gene prediction algorithms (such as honey bees), or species with few annotated orthologs in closely related species (such as *Varroa*). Another attraction is that proteogenomics can be performed with publicly available data that was originally produced for other purposes.

Whether downloaded from public repositories or produced *de novo,* expression data should ideally have organ- and tissue-level resolution, cross a range of life stages, and include samples from both sexes. This is important because tissue-specific expression is widespread: in humans, 19% of expressed proteins show tissue specificity[162] and almost 65% of detectable genes with three or more exons contain a tissue-specific splice isoform[163]. Both honey bees and *Varroa*, which are evolutionarily distant from the most common model organisms and far less studied, could benefit from annotation refinement by proteogenomics.

1.6 Hypotheses and aims

1.6.1 Research goals

My goals in this thesis are two-fold. The first goal is to improve on the existing protein databases for honey bees and *Varroa* to enable further study of these species and assist with future gene annotation efforts. Such an endeavor will not only improve future proteomics experiments on these organisms and their host-parasite interactions, it may also unlock data existing in previous proteomics datasets for new interpretation. To achieve this goal, we ask the following questions: 1) Are there novel protein-coding regions in the honey bee and *Varroa* genomes and 2) Can we construct new protein databases to learn more about these organisms' fundamental biology?

The second goal is to elucidate some of the molecular mechanisms that underlie one of honey bees' defenses against *Varroa* and brood pathogens: hygienic behaviour. Many differential expression studies comparing hygienic and non-hygienic honey bees (as well as high VSH and low VSH bees) have already been performed, but very few have identified molecules that induce hygienic behaviour and none have investigated their interactions with specific genes or proteins. Therefore, we ask the following questions: 1) Can we identify specific hygienic-behaviour stimulating odorants associated with freeze-killed and *Varroa*-infested brood that are absent or reduced in healthy brood? And 2) Do these odorants interact with hygienic behaviour-associated odorant binding proteins?

1.6.2 **Hypothesis 1**

Here, we address the first goal of improving on existing protein databases. We hypothesize that there are unannotated protein coding regions in both the honey bee and *Varroa* genomes, since this could at least partially explain the pervasively low peptide identifications. We aim to test this hypothesis using a proteogenomics approach in which we search mass spectrometry data against a six-frame genome translation, followed by orthology delineation, transcript confirmation, and a final proteomics search to determine any increases in identification rates. We package any candidate novel sequences into protein databases suitable for use in future mass spectrometry workflows and suggest that the data be included in structural annotation refinement efforts.

For honey bees, there are numerous comprehensive proteomics datasets to utilize for proteogenomics (most of which have already been incorporated into on-line honey bee peptide[164]

and protein[125] atlases); therefore, we will primarily use publicly available data for this endeavor. However, for *Varroa*, very few proteomics datasets exist. Those that do are not spectra-rich and do not cross developmental stages or sexes. Therefore, we will generate a new high-quality quantitative proteomics dataset to facilitate the proteogenomics effort. Since such a dataset is novel for *Varroa,* we will also take the opportunity to interrogate this data to identify key features of *Varroa* developmental biology and sexual differentiation. The resources we generate from this work will help inform future concerted annotation efforts and improve the quality of proteomics research on honey bees, *Varroa,* and their interactions with one another.

1.6.3 **Hypothesis 2**

Here, we address the first part of our second goal of elucidating molecular mechanisms behind hygienic behaviour: identifying hygienic behaviour-stimulating odorants. We will focus on freeze-killed brood and *Varroa* because 1) the FKB assay is an effective method for scoring colonies for hygienic behaviour and therefore FKB must emit hygienic behaviour-inducing odorants, 2) because *Varroa* is the most notorious and pervasive honey bee pest in the world, and 3) recent evidence of "social apoptosis" in the Eastern honey bee suggests that self-terminating (dead) infested brood may in part underlie VSH. We hypothesize that specific odorant cues are released from FKB and *Varroa*-infested brood, which hygienic bees are more sensitive to than non-hygienic bees. To test this, we aim to first compare odorant profiles between healthy and dead/diseased brood states using two complimentary GC-MS extraction methods: solid phase microextraction (SPME) and solvent extraction. Next, we will investigate key differentially emitted compounds *ex vivo* (using electroantennography) and *in vivo* (using behavioural assays). These are two complimentary ways of testing whether hygienic bees are more sensitive to these

odorants. Electroantennography, which measures the cumulative nerve depolarizations across an antenna upon odorant application, will allow us to compare the strength by which the odorants stimulate the antennae of hygienic and non-hygienic bees in the absence of other odorants. Behavioural assays, on the other hand, will test how these sensitivity differences may translate into a behavioural output in the context of the hive.

Despite being the main method for hygienic behaviour selection, odorants emitted from FKB have never been analyzed. This research will begin to unravel the molecular mechanism behind hygienic behaviour by identifying some of the stimulating molecular players involved.

1.6.4 **Hypothesis 3**

Finally, we will address the second part of our second goal: determining if hygienic behaviour-stimulating odorants interact with hygienic behaviour-associated OBPs. We will focus on OBP16 and OBP18 because these have the strongest statistical link to hygienic behaviour. We hypothesize that the death and *Varroa* odorants interact with OBP16 and OBP18 more strongly than odorants that are not associated with hygienic behaviour. We aim to test this by 1) comparing affinities of odorant molecules to recombinant OBP16 and OBP18 using *in vitro* ligand binding assays, 2) testing RNAi strategies to knock down OBP16 and OBP18 for eventual behavioural assays, and 3) beginning to develop a reciprocal transgenic strategy to overexpress OBP16 and OBP18.

In vivo techniques for gene manipulation are generally not well-developed in honey bees, and the antennae have never been the target of such experiments. Nevertheless, even showing physical

OBP-ligand interactions with *in vitro* experiments will be some of the first direct evidence of their involvement in detecting hygienic behaviour-inducing odorants, specifically. This information would add substantial strength to correlational observations between these proteins and hygienicity, and will help us understand how simple molecular interactions can govern a complex behaviour.

Chapter 2: Developing community-wide resources for further proteomic studies of honey bees (*A. mellifera*) and mites (*V. destructor*)[2]

2.1 Introduction and rationale

Today, developing sustainable methods of *Varroa* control is a top priority for the beekeeping industry. Social immunity traits are honey bees' main defense against this parasite, and selective breeding programs have succeeded in developing mite-resistant populations[36,112]. Other active areas of mite-control research include developing RNA interference strategies targeting essential biological processes in *Varroa*[165-168]. Despite this interest, surprisingly little is known about the fundamental molecular biology of the *Varroa* mite. Since the egg, protonymph and deutonymph life stages only exist when the foundress mite (reproductive female) is actively reproducing within capped honey bee brood comb[22], they are seldom observed and are tedious to sample (Figure 2.1). Furthermore, all male mites die soon after the adult honey bee emerges so even though they are obviously important factors in mite reproduction, our knowledge of their basic molecular biology is extremely limited. Research on *Varroa* has focused on its role as a vector for viruses[18,26,169-172], their response to pheromone cues[173-175], attempts to control it via RNAi[165-167] and host shifts[176]. Prior to this study, there were only two *Varroa* proteomic investigations, one of which focused on viral proteins[170] and the other identified fewer than 700 proteins and

[2] Content in this chapter has been published: McAfee, A. Michaud, S., and Foster, L. J. A controlled, cross-species dataset for exploring biases in genome annotation and modification profiles. *Data Brief.* 5:829-33. Copyright (2015) Authors. Permission not required for reprinting; McAfee, A. *et al*. Towards an upgraded honey bee (*Apis mellifera* L.) genome annotation using proteogenomics. *J. Proteome Res*. 15:411–21. Copyright (2016) American Chemical Society. Reprinted (adapted) with permission; McAfee, A. *et al*. A *Varroa destructor* protein atlas reveals molecular underpinnings of developmental transitions and sexual differentiation. *Mol. Cell. Proteomics*. 16:2125-37. Copyright (2017) The American Society for Biochemistry and Molecular Biology. Permission not required for reprinting

analyzed just one developmental stage[177]. Global protein expression changes associated with developmental transitions and sexual differentiation are yet unknown, but could provide key information that can be used for developing targeted miticides or understanding how the mite evades host defenses.

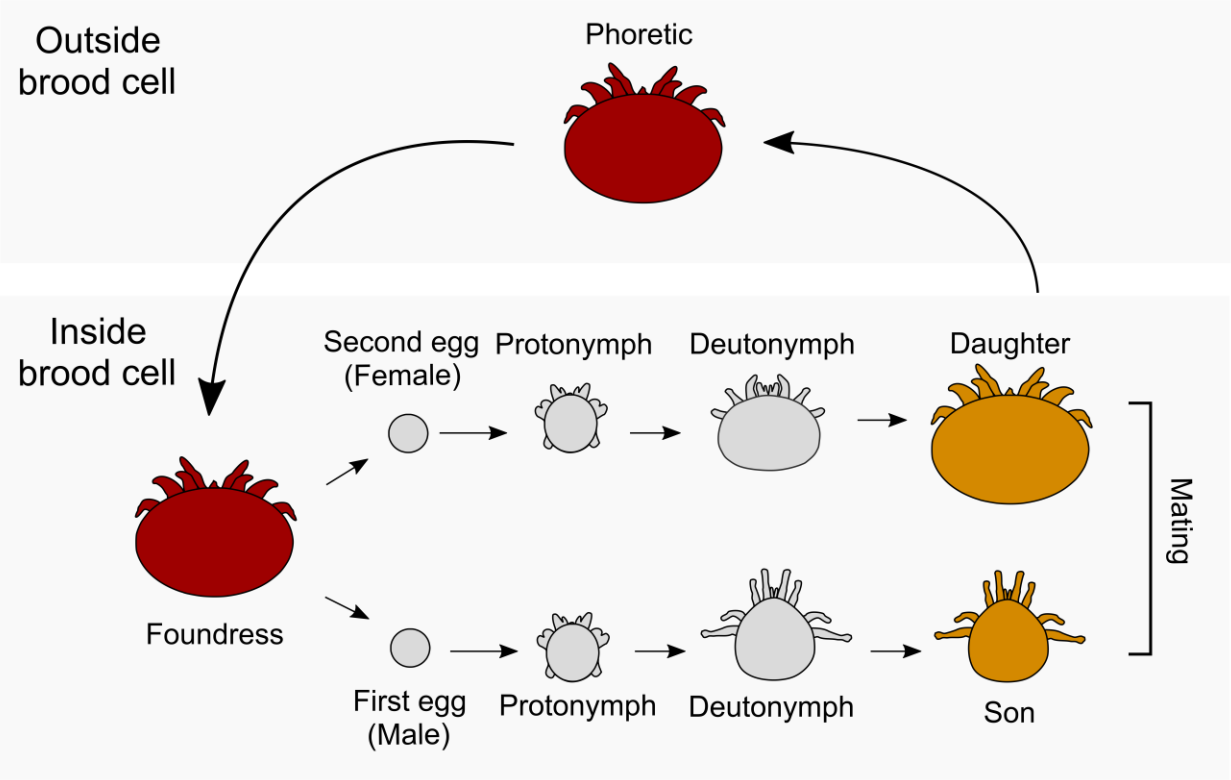


Figure 2.1. Schematic representation of the mite life cycle.

The foundress mother mite invades a honey bee brood cell just prior to capping. Soon after, she lays her first egg (haploid), which develops into a male mite. Then she lays more eggs (diploid) which develop into female mites. The adult sisters mate with their brother and emerge with the adult honey bee. The males die soon after, whereas the females enter the phretic (non-reproductive) stage and search for a new brood cell to invade. For egg and protonymph stages, males and females are visually indistinguishable. Colours indicate melanisation of the cuticle and sizes are not proportional.

The *Varroa* genome was first sequenced in 2010[21] and was accompanied by a provisional gene annotation. Gene annotations are living databases and, particularly with newly

sequenced species, they undergo continuous refinement as more 'omic data becomes available. Unfortunately, the more evolutionarily distant a species is from well-annotated species typically used for orthology delineation and gene prediction training sets, the less accurate the predictions tend to become. Such is the case for both *Varroa* and honey bees[125,151]. Proteogenomics[178,179] can help overcome this problem by sequencing the expressed protein regions in a relatively unbiased survey of the genomic landscape. Since protein expression is dynamic throughout an organism's life cycle, high resolution 'omics data that crosses developmental stages and sexes is very well-suited for this purpose and maximizes the chance of detecting expressed genes. Appropriate data for proteogenomics already exists for honey bees in publicly available repositories; however, for *Varroa*, such data has not yet been generated. Both the honey bee and *Varroa* annotations could have many yet-unannotated protein-coding regions in their genomes.

2.2 Research goals

Our goal is to improve upon existing gene annotations in both *Varroa* and honey bees, and in the process produce global *Varroa* protein expression profiles throughout development of both sexes to provide a foundational understanding of *Varroa* biology. Our first aim is to conduct a quantitative proteomics investigation crossing all major developmental stages (egg, protonymph, deutonymph, adult) of both males and females, where distinguishable. Our second aim is to use this data to conduct a proteogenomic investigation for *Varroa* and to use previously existing proteomics data to do the same for honey bees. Our third aim is to use the *Varroa* developmental data for a differential expression analysis between

developmental stages and sexes, while providing a few examples of biological inquiries and insights it can be used for. Finally, to turn this *Varroa* protein abundance data into a community-wide resource, we present the information as an interactive, web-based protein atlas (http://foster.nce.ubc.ca/varroa/index.html), enabling further hypothesis generation regarding *Varroa* biology and *Varroa*-honey bee interactions.

2.3 Developing a more comprehensive *Varroa* genome annotation

Procuring an accurate protein database is critically important for proteomics applications. An updated *Varroa* genome assembly was recently released (ADDG00000000.2) and new gene set will soon to be released, which we have made provisionally available through ProteomeXchange (PXD006072). To test the accuracy of the new gene set compared to the first draft (published in 2010), we searched our complete *Varroa* proteomics data against both versions and found that greater than 2-fold more unique peptides were identified using the refined annotation (Figure 2.2 A). Overall, we identified nearly 20,000 unique peptides corresponding to 3,102 protein groups at 1% peptide and protein FDR (Figure 2.2 B**)** representing the first global survey of *Varroa* protein expression across sexes and developmental stages.

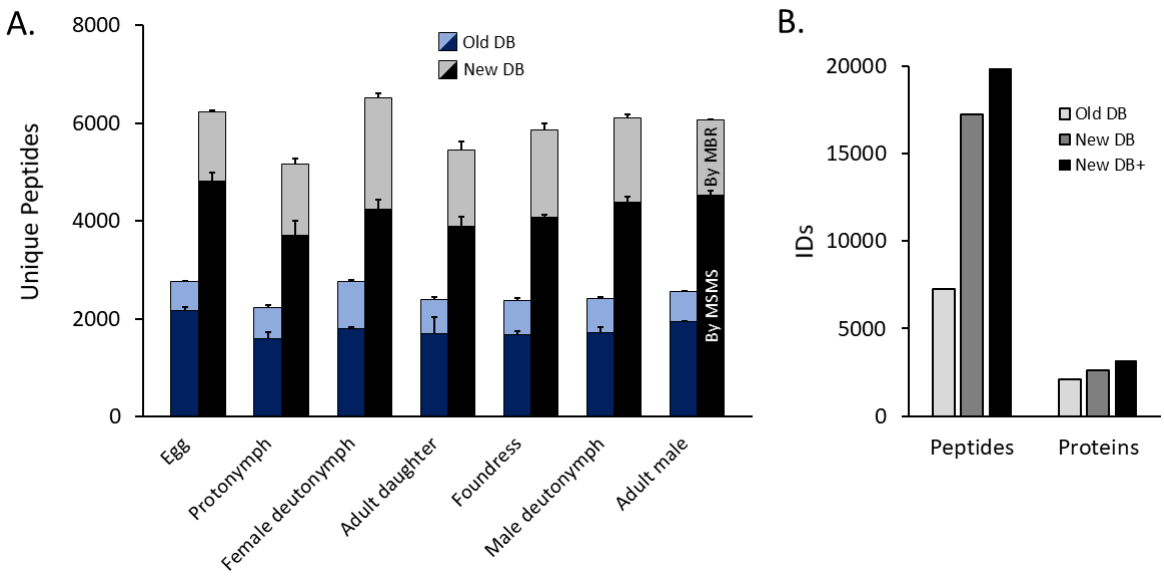


Figure 2.2. Overall *Varroa* peptide and protein identifications crossing developmental stages.

MSMS data was searched against the initial, provisional draft *Varroa* gene annotation ("Old DB") and the most recent updated annotation ("New DB", unpublished, supplied by Jay Evans). The data included biological triplicates of each developmental stage and all protein databases also included NCBI *Varroa* sequences and all viruses known to infect honey bees and *Varroa*. A. Light stacks represent peptide identifications via match between runs (MBR) and dark stacks represent identifications via MSMS matching. Error bars are standard deviation. B. Cumulative identifications. "New DB+" refers to the most recent annotation plus all honey bee proteins and new fragments identified by our proteogenomics effort.

2.3.1 Proteogenomics to identify unannotated protein-coding regions in *Varroa*

2.3.1.1 Candidate protein-coding regions in *Varroa*

Despite a dramatic improvement in accuracy over the initial draft annotation, the current annotation could likely be further improved through proteogenomics. We searched the MSMS data against a six-frame genome translation database and identified 519 new candidate protein-coding regions at 1% FDR which were absent from the current annotation. Three hundred and one of these were supported by two or more peptides. Furthermore, 169 of these protein groups were differentially expressed through

development (Figure 2.3 A). The Human Proteome Organization has published guidelines for extraordinary detection claims (such as missing proteins), which include providing annotated spectra, considering alternate explanations of peptide spectral matches, using synthetic peptides to validate detection claims, and supporting novel protein sequences with two or more independent peptides[158]. While this approach is necessary for a highly scrutinized genomes, such as for humans, we did not apply all these filters to our data. Our rationale is that if the intent is to integrate the peptide or protein data into informant datasets for future iterations of structural annotation, this adds an extra layer of false discovery control. In fact, the informant gene set is normally composed of suggested expressed regions (often from different species altogether). The new gene predictions that may result from using such data, however, should still be manually curated and validated. In the meantime, these candidate novel sequences were combined with previously annotated sequences into a single fasta file for future mass spectrometry data analyses and to facilitate integration into informant gene sets. The database as well as all raw mass spectrometry data files, are available on the ProteomeXchange public repository (www.proteomexchange.org; PXD006072). Annotated spectra are available at MS-viewer (search key: msmx6z444s).

2.3.1.2 Why did algorithms fail to predict these candidate expressed regions?

Since missed genes appear to be a common problem in genome annotation, we sought to investigate the root cause of failing to locate these sequences in the first place. Gene prediction algorithms often use training gene sets from well-annotated species with similar

genomic properties to help define genes in the newly sequenced target species[178]. One reason why an algorithm might fail to identify expressed sequences is if they occur in regions with significantly different AT content or codon bias (indeed, this is precisely what happened during the honey bee annotation[9]), so we compared these properties between the newly identified protein coding regions and the previously known coding regions identified in the same six-frame translation search. We found that the newly identified regions had the same AT content as the previously known regions, which were both significantly different from the genomic average (Figure 2.3 B). While this lends additional confidence that the proposed new coding regions do not occur by chance, it does not explain why they were missed in the annotation. Furthermore, the amino acid composition and nucleotide positional codon bias (Figure 2.4) was the same between the new and known coding regions.

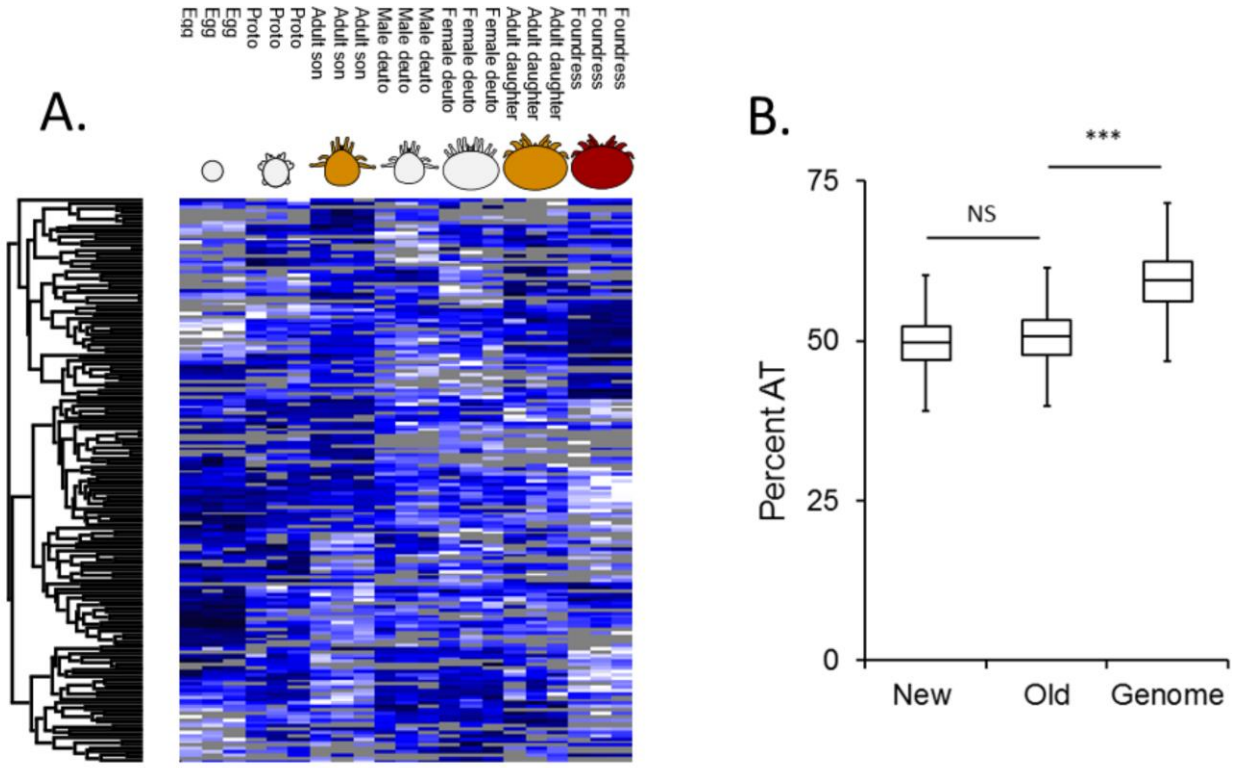


Figure 2.3. A search against a six-frame genome translation database proposes new protein-coding fragments.

A. New candidate protein-coding fragments which are differentially regulated across development. Grey tiles represent missing data, white tiles indicate low Z-score, and dark blue tiles indicate high Z-score. Hierarchical clustering was performed in Perseus using average Euclidian distance (300 clusters, maximum 10 iterations). Statistics were performed using an ANOVA (Benjamini Hochberg-corrected FDR = 5%). B. Comparison of the AT composition between newly identified candidate sequences, previously known sequences (old) and the *in silico* fragmented genome. Statistics were performed using a one-way ANOVA (3 levels) and a Tukey HSD post-hoc test. NS: not significant. * indicates $p < 0.0001$. Boxes depict the interquartile range (IQR) and whiskers span 1.5*IQR.**

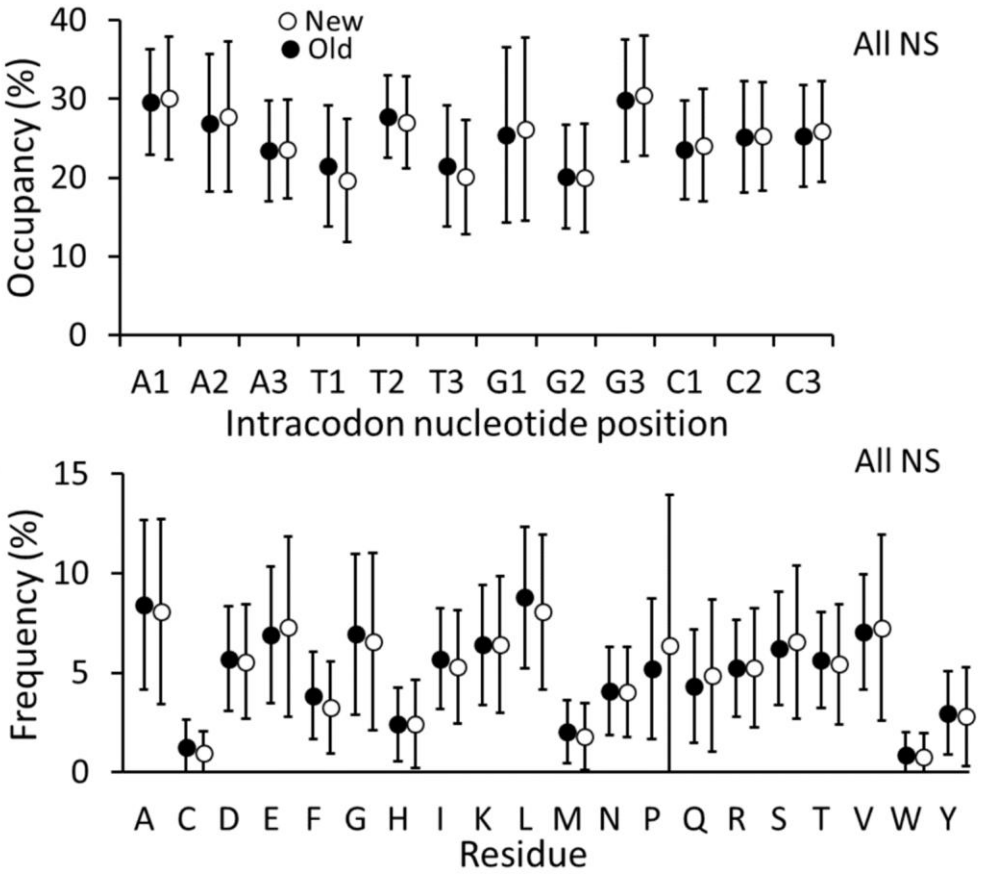


Figure 2.4. Comparison of nucleotide and amino acid residue frequencies between new candidate sequences and old (previously annotated) *Varroa* sequences.

All comparisons were not significant (NS) for mean amino acid composition or intracodon nucleotide position of new protein coding regions compared to old (one-way ANOVA). Error bars represent standard deviation.

Since some algorithms rely on homology evidence to support annotations, one reason sequences may not be annotated is if they do not have known orthologs. We used Blast2GO to identify potential orthologs and found that nearly 72% (377) of the sequences had significant similarity (e-value cut-off: 1E-5) to at least one sequence in the non-redundant NCBI protein database (Figure 2.5). Of those, the majority (85%) matched to sequences from other members of phylum Arthropoda but some significant Chordata, Nematoda, Mollusca and Annelida matches were also present. Only nine sequences significantly

matched well-annotated species (*Homo sapiens*, *Mus musculus*, *Drosophila melanogaster* and *Caenorhabditis elegans*) and 148 (28%) had no significant sequence similarity to any species, despite 54 of them being supported by two or more peptides. In addition, 5 sequences were highly similar to known honey bee sequences, suggesting these are likely the result of DNA contamination within the *Varroa* sample used for DNA sequencing. This is not surprising since honey bee tissue is the mite's sole food source, so some contamination of this nature is expected. We removed these sequences since we include all honey bee proteins in our search database regardless in order to account for abundant honey bee proteins consumed by *Varroa*. All other fragments identified through proteogenomics were added to the protein database and utilized in subsequent analyses.

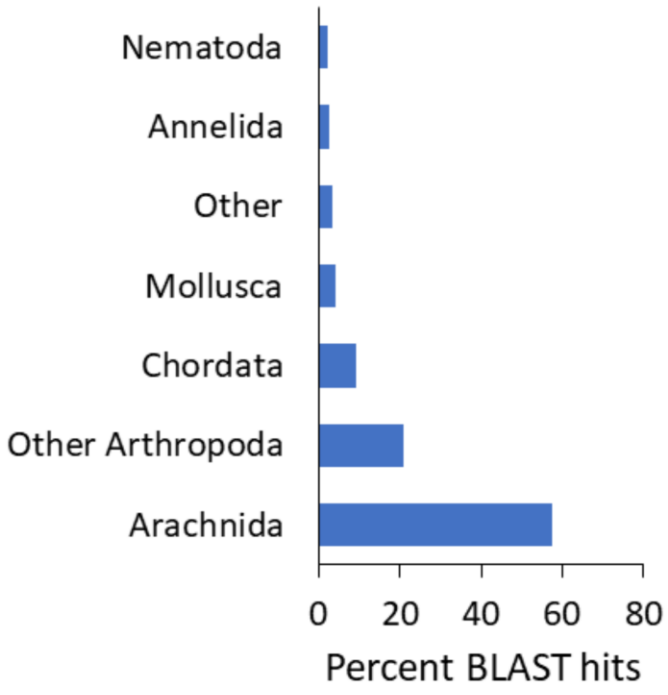


Figure 2.5. BLAST sequence alignment summary of the new protein sequences for major (> 1% frequency) taxa.

2.4 Developing a more comprehensive honey bee genome annotation

The honey bee genome has been sequenced and annotated twice over, enabling proteomics and functional genomics methods for probing relevant aspects of honey bee biology. Unlike *Varroa*, by now there have been numerous large-scale proteomic interrogations of the honey bee, but one troubling trend that emerged from proteomic analyses is that honey bee samples consistently result in lower peptide identification rates compared to other organisms. This suggests that either the genome annotation can be improved or that atypical biological processes are interfering with the mass spectrometry workflow, or both. We and others have published several proteomic analyses of honey bees where we have observed that, anecdotally, fewer honey bee proteins are identified than we would expect[125,180] compared to similar analyses on other systems[181-183].

2.4.1 Deficiencies in honey bee proteomics compared to commonly studied species

Compared to similar large-scale proteomic studies in other species, the honey bee has fewer observed proteins (Fig 2.6 A). These experiments are not directly comparable because they are from different laboratories with different sample preparation and LC-MS methods. However, even within our own group we have been able to routinely identify over twice the number of proteins in human THP-1 cells[183] (4,997 proteins, 1% FDR, 24% proteome coverage) compared to the most that has ever been previously reported in honey bees[125] (2,288, 1.2% FDR, 15% proteome coverage).

Here, we further support this observation by comparing protein identifications across species (*S. cerevisiae, D. melanogaster, C. elegans, M. musculus,* and *H. sapiens*) while controlling as many extraneous variables as possible (Figure 2.6 B). While it is possible that protein extraction

efficiency could vary with different protein sources (*e.g.* cell culture vs. liver), all sample preparation steps downstream of extraction were identical. As predicted from anecdotal observations, the results show that significantly fewer proteins were identified in honey bees compared to all of the other species tested (T-tests, $p < 0.005$, Bonferroni-corrected $p < 0.03$). Furthermore, when protein identifications were normalized to the size of the proteome, honey bees still had the lowest proteome coverage (Figure 2.6 C). Together with the literature comparison of maximum protein identifications between species, these observations suggest that even using standardized proteomic techniques, it is difficult to achieve a high level of proteome coverage in honey bees.

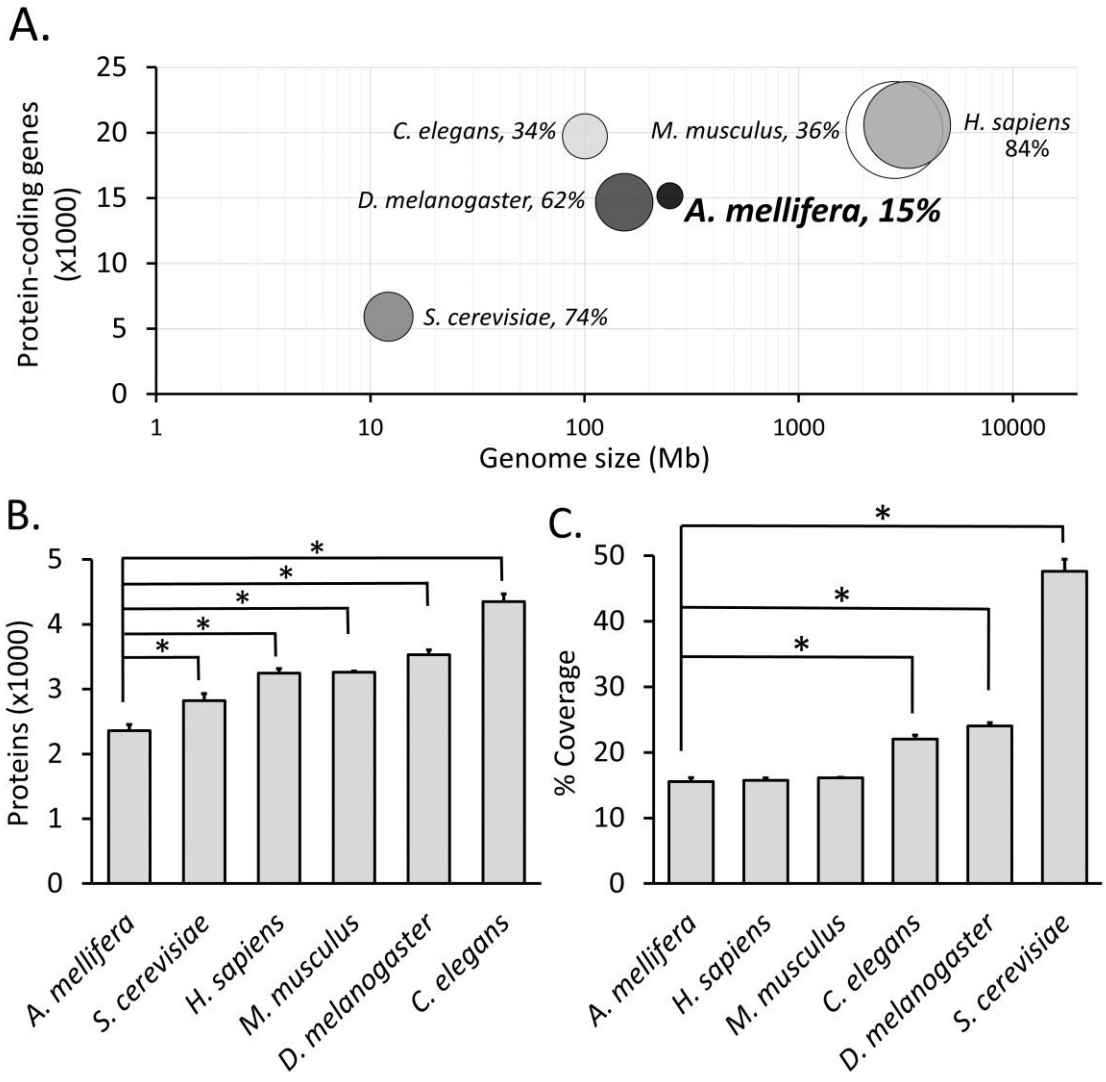


Figure 2.6. Comparison of proteomes and genomes across species.

A. **The number protein coding genes were obtained from Wilhelm *et al.* (2014)[152], Brunner *et al.* (2007)[155], Hillier *et al.* (2005)[184], Church *et al.* (2009)[185], www.beebase.org and www.yeastgenome.org. Genome size was determined by the most recent or the most complete genome sequence published on NCBI. Circle size represents the number of amino acid residues in the database, and the percent represents the fraction of the proteome demonstrated to be accessible by mass spectrometry. B. A controlled proteomics experiment comparing protein identifications across species. Samples comprised of *Apis melliera* (5th instar worker larvae), *Saccharomyces cerevisiae*, *Homo sapiens* (HeLa cells), *Mus musculus* (liver), *Drosophila melanogaster* (adult females), and *Caenorhabditis elegans* (mixed life stages). The number of proteins identified in *A. mellifera* samples was compared to every other species using T-tests. Asterisks indicate statistical significance ($p < 0.005$; Bonferroni-corrected $p < 0.03$). C. Data was normalized to the number of protein-coding genes. Asterisks indicate statistical significance. Percent coverage of the *A. mellifera* proteome was compared to all other species using T-tests ($p < 0.0002$; Bonferroni-corrected $p < 0.001$).**

2.4.2 **Proteogenomics to identify unannotated protein-coding regions in honey bees**

One explanation for generally poor honey bee peptide and protein identification rates is that there could be many unannotated protein-coding sequences. We therefore applied a systematic approach to identify missing sequences and propose new candidate ORFs.

2.4.2.1 Sequences discarded from OGSv1.0

The upgrade from the old honey bee gene annotation (OGSv1.0; 10,157 entries) to the new annotation (OGSv3.2; 15,314 entries) occurred through a re-modelling process that was aimed at improving the quality of the database. Surprisingly, only 4,083 protein sequences and 48% of tryptic peptides were fully retained in the upgraded version. Furthermore, since producing the OGSv3.2 annotation relied more heavily on transcriptome rather than proteome data[9], and since OGSv3.2 was produced independently of OGSv1.0 rather than being built upon it, we speculate that the loss of thousands of sequences from the database was a result of insufficient lines of evidence rather than active exclusion. To investigate if there are discarded proteins that should be revived, we searched the same comprehensive honey bee proteomics data set described above against a composite OGS that included all unique sequences in both OGSv1.0 and OGSv3.2. By using a composite database, we ensure that any spectra that match to OGSv1.0 do not have a higher scoring match in OGSv3.2. We identified a total of 2,612 unique peptides (746 protein groups) that were specific to OGSv1.0 (Table 2.1; 1% FDR) out of a total of 30,916 identified peptides overall. In other words, 8% of all identified peptides originated from sequences that no longer exist in OGSv3.2. While RNA-seq data has great utility as a genome annotation tool, confirmation at the protein level, as we have shown here, is the ultimate test for expression of

predicted protein-coding genes. We therefore recommend that these sequences should be re-entered into the next OGS version.

2.4.2.2 Newly identified candidate protein-coding regions in honey bees

Genes could be missed during annotation for a variety of reasons, including being mis-assigned as ncRNA, lack of homology to genes in other organisms, or being shorter than a minimum threshold. In addition, alternative splicing could produce isoforms with segments of unannotated coding regions within known genes. To gauge the potential for uncovering unknown coding regions in the honey bee, we compared the genome annotations across model organisms and found that honey bees have fewer protein sequences than would be expected based on the size of their genome (Figure 2.6 A), supporting the hypothesis that there could be a large number of coding regions missing in OGSv3.2. As a first step towards refining the genome annotation, we searched all 1,594 raw files that were used to produce the Honey Bee Peptide Atlas[164] against a six-frame translation of the entire genome (216 million residues). We identified 1,426 sequences from this search that are supported by two or more unique peptides (1% PSM, peptide, and protein FDR) but that are not present in OGSv3.2. As an illustration, three sequences with high coverage of candidate novel peptides (Contig4.16, starting nucleotide: 480165, direction: + (coded within the access as: 4.16:480165+), 8.11:336226+ and Un2727:504+) are shown in Figure 2.7 A. We confirmed the expression of these sequences by RT-PCR (Figure 2.7 B) from both adult and larval honey bee tissue, demonstrating that at least some of the candidate protein-coding regions are expressed by an orthologous method. These sequences are good candidates to include in future annotation informant "gene sets."

As with the *Varroa* proteogenomics effort, these sequences do not satisfy all the criteria suggested for extraordinary detection claims in human proteomics samples[158]. We are not outright asserting that these are "new genes." Rather, we are using this investigation to a) provide a preliminary interrogation as to whether missing gene annotations could explain the low peptide identifications we observe and b) provide additional informant data for future annotations. At this stage, we aim to be as inclusive as possible since the benefits of including additional true positive informant sequences likely outweigh the drawbacks of also including some fraction of false positives. Annotation outputs can be further curated, so any false positive structural annotations that are generated could be identified and removed.

Table 2.1. Identification of coding sequences missing from the current genome annotation.

Sequence source	Description	Protein groups
ncRNA	13 unique peptides	9
Genome translation	2+ peptides/candidate protein*	1426
	1 peptide/candidate protein*‡	488
	2+ peptides + BLAST support	194
	1 peptide + BLAST support	299
OGSv1.0	1+ peptides/candidate protein	748
Total non-redundant sequences		2546

* These sequences did not show significant sequence similarity to any other species.
‡ These sequences were not included in subsequent analyses.

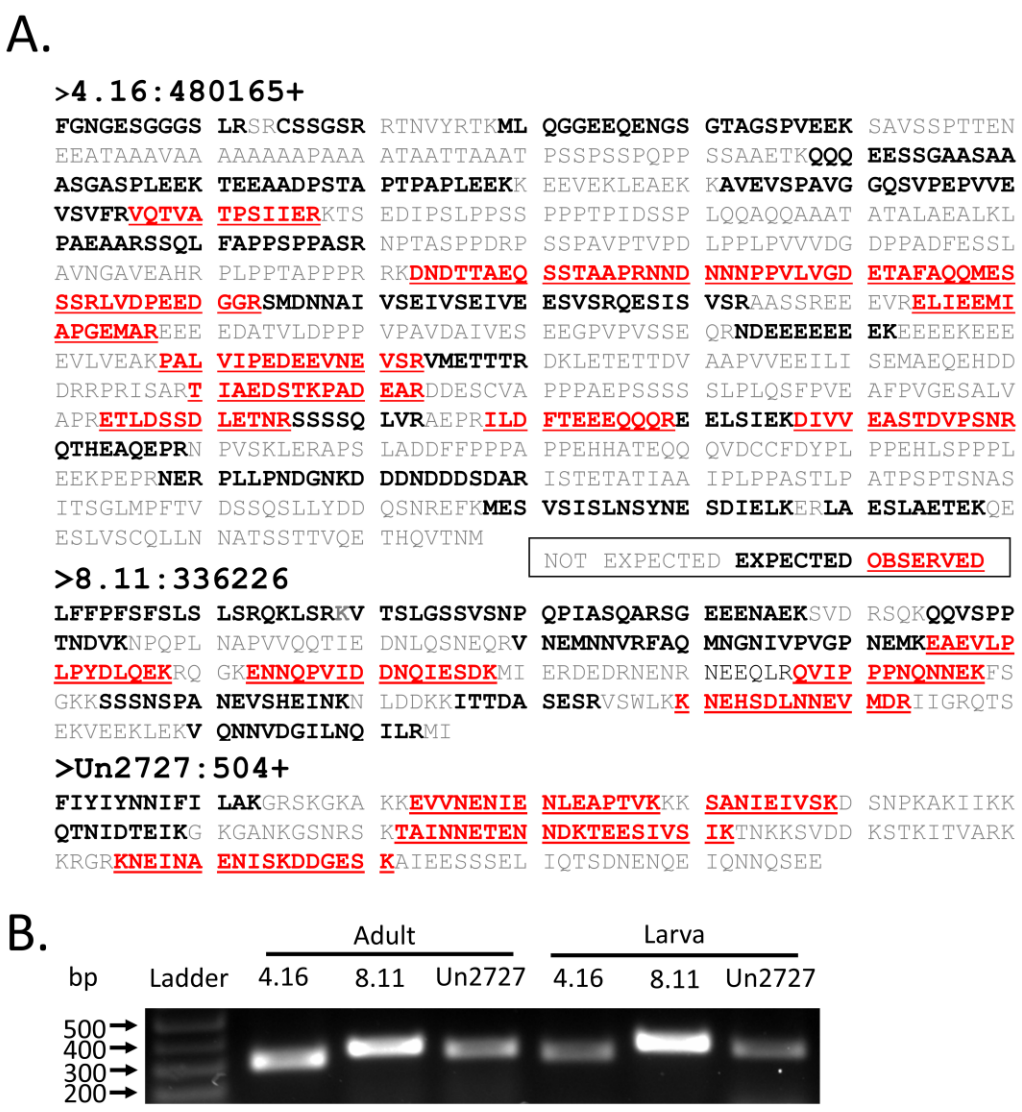


Figure 2.7. Examples of three novel honey bee protein-coding regions with high peptide coverage and confirmation by RT-PCR.

A. Accessions are in the format: contig:start position:strand. Grey regions represent tryptic peptides that are not expected to be accessable by typical mass spectrometry proteomics based on their amino acid sequence length alone (< 6 or > 26 residues). Likewise, black regions are expected to be accessable and underlined regions were actually observed in the honey bee Peptide Atlas data when searched against a 6-frame genome translation database (MaxQuant, 1% FDR). B. Nucleotide sequences spanning at least two peptides in A that were not included in the official gene set were amplified by RT-PCR.

To further explore the newly discovered proteins, we used BLAST to identify which of the corresponding nucleotide sequences have significant similarity to those found in other species (Table 2.2). For virtually all of the new candidate coding regions with significant similarity, the best BLAST hit was from another member of the Hymenoptera, even when searched against a background of the entire non-redundant nucleotide collection. Many of these also matched to honey bee sequences that are annotated as purely bioinformatics predictions in NCBI (coded by "XP" accession numbers) that have no recorded experimental support. A small fraction of new coding regions returned significant similarity to species of bacteria, which could represent honey bee commensals or other bacteria picked up while foraging. The subset of coding regions (197 in total) that are supported by two peptides as well as BLAST sequence similarity are highly confident identifications and are good targets for functional studies in the future.

We then sequenced each of the coding regions that were subjected to RT-PCR above to explore why they might have been missed during genome annotation. Closer inspection of sequence 4.16:480165+ shows that it has significant similarity (E-value = 2.00×10^{-21}) to the predicted honey bee protein sequence XP_392304.4 so it is likely that this was rejected from OGSv3.2 simply because there was no previous experimental evidence supporting its expression. Sequencing the region corresponding to 8.11:336226+, however, revealed that it is clearly an alternative splice isoform of the OGSv3.2 sequence GB52827 (Figure 2.8). ClustalW alignment of the 8.11:336226+, GB52827 and the genomic sequence shows that 8.11: 336226+ shares at least two splice sites with GB52827 (AAAGTA and CAGAAA); however, 8.11: 336226+ is not spliced at the downstream site AAAAAT whereas GB52827 is. Finally, Un2727:504+ shows significant similarity to predicted sequences in other hymenopterans (*Apis*, *Bombus* and

Megachile), but not to *Apis mellifera* specifically, and to our knowledge this is the first reported expression of the new candidate. Even though it has not been previously predicted, here we have accumulated evidence at the peptide level (MSMS data), transcript level (RT-PCR) and evolutionary level (BLAST), making a very strong case that it is a real protein-coding region.

```
8.11_RT-PCR       AAAAATAACGAGGACACGGATAGCACGATTGAACATTTGAGGAAGGAGGTGATTCAATTG
GB52827           AAAAATAACGAGGACACGGATAGCACGATTGAACATTTGAGGAAGGAGGTGATTCAATTG
Genomic           AAAAATAACGAGGACACGGATAGCACGATTGAACATTTGAGGAAGGAGGTGATTCAATTG
8.11_RT-PCR       AAGCGCGAGCTGGACGAAGCTAAA-----------------------------------
GB52827           AAGCGCGAGCTGGACGAAGCTAAA-----------------------------------
Genomic           AAGCGCGAGCTGGACGAAGCTAAAGTAAATCAATTTTTCTCTATTACATGAATAATAATA
8.11_RT-PCR       ------------------------------------------------------------
GB52827           ------------------------------------------------------------
Genomic           ATAATAAAAAAAAAAAATAATCCTGATTAATTAATTTCTAAGAGATTGATTGACGGTATAA
8.11_RT-PCR       ---------------------------------------------AAATTATCACGCA
GB52827           ---------------------------------------------AAATTATCACGCA
Genomic           CCTAACTTTTTTTTCCTTTCTCTTTCTCTCTCTCTCTCTCTCGGCAGAAATTATCACGCA
8.11_RT-PCR       GAGTTACTTCTCTTGGATCAAGCGTATCGAATCCGCAACCAATAGCCTCACAAGCGAGAT
GB52827           AAGTTACTTCTCTTGGATCAAGCGTATCGAATCCGCAACCAATAGCCTCACAAGCGAGAT
Genomic           AAGTTACTTCTCTTGGATCAAGCGTATCGAATCCGCAACCAATAGCCTCACAAGCGAGAT
8.11_RT-PCR       CTGGGGAGGAAGAGAACGCAGAAAAATCTGTTGATCGGTCGCAAAAACAGCAGGTAAGTC
GB52827           CTGGGGAGGAAGAGAACGCAGAAAAATCTGTTGATCGGTCGCAAAAACAGCAGGTAAGTC
Genomic           CTGGGGAGGAAGAGAACGCAGAAAAATCTGTTGATCGGTCGCAAAAACAGCAGGTAAGTC
8.11_RT-PCR       CACCGACAAATGACGTTAAAAATCCGCAGCCATTGAACGCGCCAGTCGTTTCAACAAACG
GB52827           CACCGACAAATGACGTTAAA-------------------------------------
Genomic           CACCGACAAATGACGTTAAAAATCCGCAGCCATTGAACGCGCCAGTCG-TTCAACAAACG
8.11_RT-PCR       ATTGAGGATAATCTGCAATCGAACGAGCAACGAGTGAACGAGATGAACAACGTTCGATTC
GB52827           ------------------------------------------------------------
Genomic           ATTGAGGATAATCTGCAATCGAACGAGCAACGAGTGAACGAGATGAACAACGTTCGATTC
```

Figure 2.8. Sequence alignment of the region encoded by 8.11:336226+.

Sanger sequencing was performed on the RT-PCR amplicon (top) and the sequence was aligned to the cross-referenced OGSv3.2 gene sequence (middle) as well as the genomic sequence (bottom) using ClustalW. OGSv3.2 contains only a single isoform of this gene.

To provide deeper evidence supporting expression the new candidate coding regions at the transcript level, we analyzed a publicly available RNA-seq data set (GSE61253) including mRNA samples from drone, queen and worker honey bees. We found that 574 (33%) of the candidate protein coding regions overlapped at least partially with assembled transcripts. Full overlap between the proteins and transcripts was rare, probably because the protein prediction algorithm does not recognize splice sites, whereas transcript alignments span multiple exons[186]. We did not expect all the protein sequences to be supported because although the RNA-seq data

set includes all three castes, it represents only one life stage (L5 larvae) while the proteomics data was generated from all publicly available datasets at the time (mostly adult worker). In light of this, finding support for one third of our predicted sequences is substantial.

Table 2.2. BLAST best hits for three high-coverage novel coding regions.

Accession	Species	Family	E-value
	4.16: 480165+		
XP_003486676.1	*Bombus impatiens*	Apidae	9.00E-87
XP_003402461.1	*Bombus terrestris*	Apidae	1.00E-85
XP_392304.4	*Apis mellifera*	Apidae	2.00E-18
EFN65736.1	*Camponotus floridanus*	Formicidae	7.00E-14
EZA51051.1	*Cerapachys biroi*	Formicidae	1.00E-04
EFZ14323.1	*Solenopsis invicta*	Formicidae	2.00E-04
	8.11: 336226+		
XP_003402714.1	*Bombus terrestris*	Apidae	9.00E-176
XP_003694201.1	*Apis florea*	Apidae	2.00E-135
XP_003702546.1	*Megachile rotundata*	Apidae	2.00E-23
	UA2727:504+		
XP_00660774A15:A190.1	*Apis dorsata*	Apidae	2.00E-87
XP_003698993.1	*Apis florea*	Apidae	3.00E-47
XP_003397193.1	*Bombus terrestris*	Apidae	3.00E-23
XP_003488031.1	*Bombus impatiens*	Apidae	2.00E-17
XP_003703674.1	*Megachile rotundata*	Apidae	8.00E-05

2.4.2.3 Improvements to overall peptide identification rates in honey bees

We next sought to quantify the benefit to proteomics experiments that could be gained when the 'missing coding regions' accumulated here are included in the search database. We categorized the new candidate coding regions according to the number of peptides identified per sequence, BLAST sequence similarity to other species and whether or not the sequences have been included in a previous OGS (see Table 2.1; those with only one peptide and no significant

BLAST hits were discarded). The categories are based on whether they were supported by one, two or three of these lines of evidence, and are encoded in the FASTA unique identifier. This list was added to OGSv3.2 to create a new comprehensive protein database (Figure 2.9 A), and the accuracy of this database was evaluated with respect to OGSv3.2 and OGSv1.0 using an LC-MSMS data set that was not involved in upstream identification of the new sequences themselves. As expected, the number of peptides identified increased compared to both OGSv3.2 and OGSv1.0 (Figure 2.9 B), but the increase was modest compared to what we were expecting. However, the data set used to initially find the new coding regions was very rich, with organ-level resolution and all major castes in addition to peptide fractionation, whereas the data set used to test identification improvement included only the worker caste and three tissues. It is possible that many of the new candidate coding regions described in Table 2.1 are low abundance or tissue-specific, leading them to only be observable with extensive dissection and fractionation, whereas only a fraction of them are high abundance and expressed across tissues. This could also explain in part why the sequences had not been previously annotated or retained. In the current analysis, we are unable to distinguish between sequences not observed for this reason, and false positives. Our goal was simply to compile a comprehensive list of possible protein-coding regions, not to provide thorough validation of each one. Without approaching annotation refinement with inclusivity, we may never identify the *Varroa* and honey bee genes we have missed. We have made the protein database that resulted from this work openly available[151] in order to support future honey bee proteomics studies.

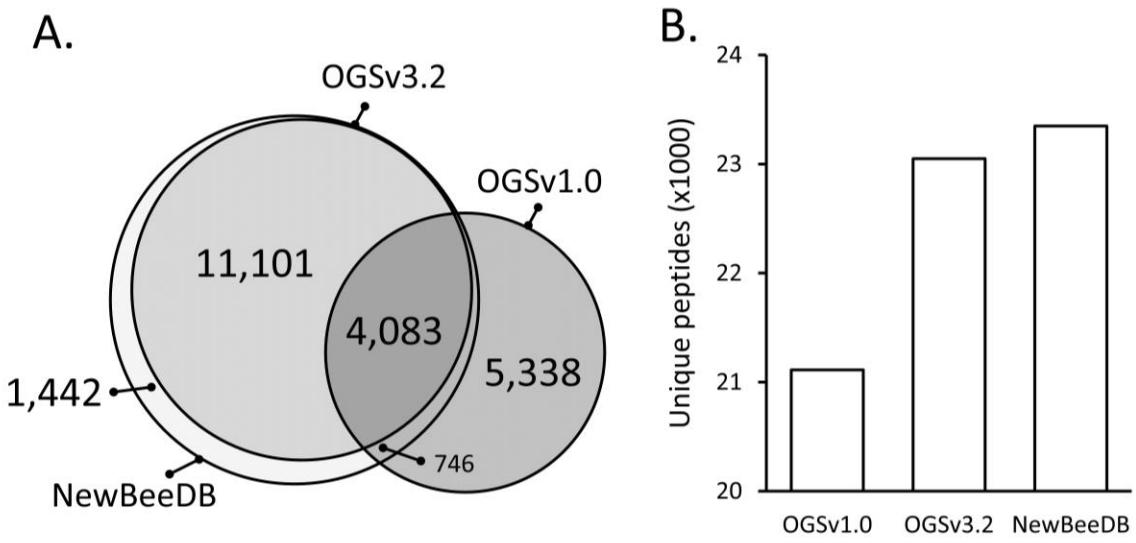


Figure 2.9. Improvements in identified spectra using a new honey bee protein database.

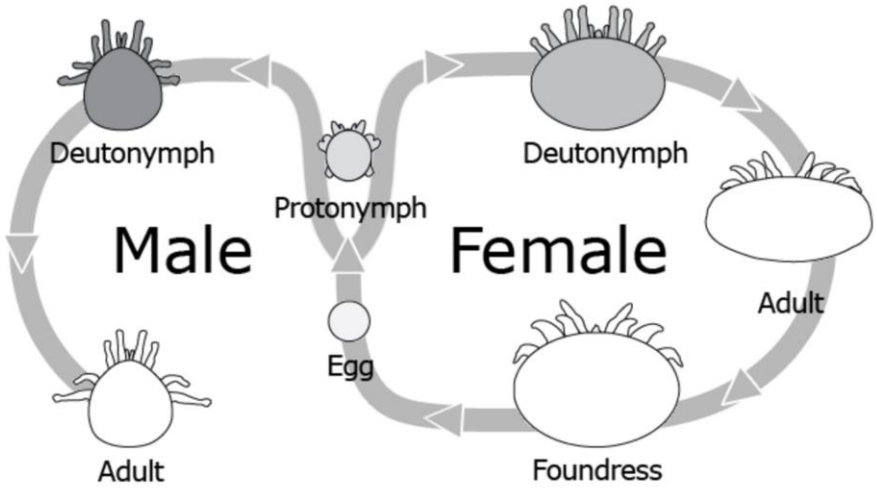
A. Protein sequence overlap between OGSv3.2 (15,314 entries), OGSv1.0 (10,157 entries) and NewBeeDB (17,372 entries). NewBeeDB was produced by adding all new non-redundant candidate sequences that are supported at the peptide level to OGSv3.2 (outlined in Table 2.1), in addition to discarded OGSv1.0 sequences for which we found peptide expression evidence. B. Honey bee proteomics data that was not used to identify new coding regions was searched against OGSv1.0, OGSv3.2 and NewBeeDB databases to evaluate improvements in identification rates. MaxQuant v1.5.2.8 was used for all searches and results were filtered at 1% protein and peptide FDR. Total unique peptides across the three replicates are shown. Note the y-axis does not begin at zero.

2.5 Insights into *Varroa* developmental biology

To maximize the utility of the information contained in the new *Varroa* proteomics dataset for researchers, in addition to the proteogenomics analysis, we incorporated all the quantified proteins into an interactive *Varroa* protein atlas (http://foster.nce.ubc.ca/varroa/index.html). The atlas features a database of the quantified proteins as well as a visual and numerical display of their relative expression of each protein in different developmental stages (Figure 2.10).

Stages: Egg Protonymph Deutonymph Male | Female Adult Male | Female Foundress Search

gi|283671561|gb|ADDG01093290.1|:True:22



The above diagram is generated from the following data:
(Varroa diagram is coloured according to percent expression, relative to the all 7 tested life stage and sex: 100% = black, 0% = white)

p<0.05	Egg	Protonymph	Deutonymph		Adult		Foundress
			Male	Female	Male	Female	
*	5	15	51	29	0	0	0

An asterisk (*) denote significance(p<0.05).

Sequence:

ACTRCPMHQDPAHVVPIQHELAHVVPVQHEPEPAVPAQHHHHRQQHHRHVQTHPVGAAPVHPHGATH
AQAVSTHGVRPAAVQQIQHVQHRIVHAEPVHAAPVVSVHQEAVAHAGPVHGAVPISHEVVHGHSGHAPI
GHAHGGIVHQQYFTIHHTQSEKKLVHAPVVTAVPGTKHDAAVDTHPVVPVRGGHGEPAVHAGPVVPVV
SHVPPHPAPLDEKETPAASESVEPITPNYGRKKRRFHLKKKSSKAASSAASASASELKPTVMDAVEEKENEQ
KED

Figure 2.10. Example page of the web-based *Varroa* protein atlas.

The atlas was constructed using the framework described for the honey bee protein atlas. Shading of the cartoon mites indicates relative expression and an asterisk indicates that this protein was significantly differentially expressed according to developmental stage. Website: http://foster.nce.ubc.ca/varroa/index.html.

2.5.1 **Vitellogenin expression throughout development**

Of the 3,102 proteins identified, 1,433 were differentially expressed across developmental stages (5% Benjamini-Hochberg corrected FDR; Figure 2.11 A, Table 2.3). As a quality control method, we specifically analyzed vitellogenin (an evolutionarily conserved yolk protein) expression since this is one of the only proteins where the developmental patterns of expression are known[187]. We expected to see high levels of vitellogenin-1 and vitellogenin-2 in the foundress and egg, with quantities decreasing approaching adulthood and indeed, this is what was observed (Figure 2.11 B). Interestingly, some of the novel peptides identified in our proteogenomic effort mapped back to protein fragments with significant sequence similarity to vitellogenin, and upon closer inspection we found that some of these peptides are simply non-synonymous single nucleotide sequence variants of this well-known gene. However, we also identified novel protein fragments with significant similarity to vitellogenin that did not physically overlap with the known vitellogenin genes (Figure 2.11 C). Like vitellogenin-1 and 2, the highest protein abundance for these candidate novel sequences was in the egg. Furthermore, they group into two clusters of expressed fragments (one two-fragment cluster and one four-fragment cluster) closely linked on two different contigs, suggesting that the fragments form exons of two different genes (Figure 2.11 D) and clearly illustrate how mass spectrometry data can aid in gene predictions.

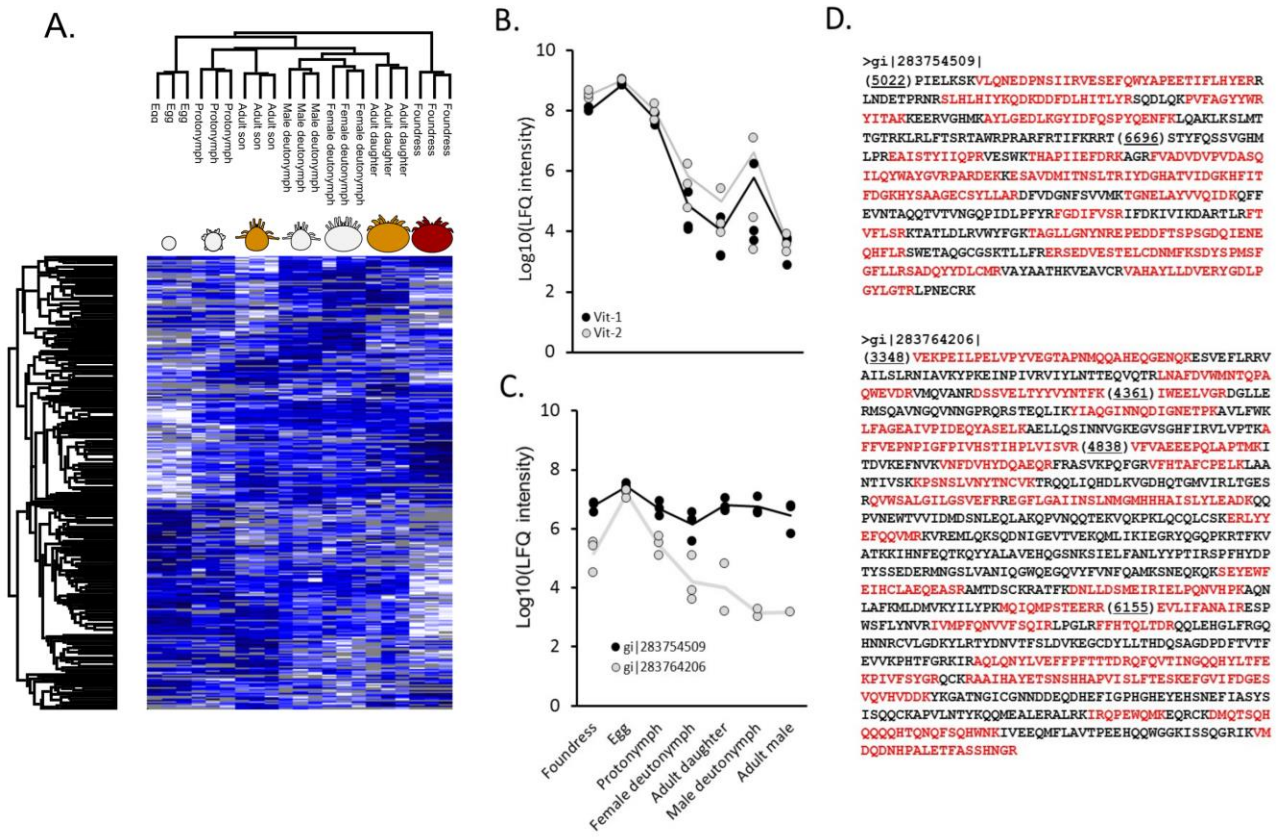


Figure 2.11. Analysis of vitellogenin expression across *Varroa* life stages.

A. Heatmap showing significantly differentially expressed proteins across developmental stages (ANOVA; Benjamini Hochberg-corrected FDR = 5%). Grey tiles indicate missing data. Proto and deuto refer to protonymph and deutonymph, respectively. B. Vitellogenin (Vit)-1 and Vitellogenin-2 protein expression across developmental stages. C. Expression of new protein fragments showing significant vitellogenin homology. D. New vitellogenin protein fragment sequences. Observed peptides are red, the fasta header indicates the contig number and bracketed numbers indicate the nucleotide start position of each fragment within the contig. Both protein sequences were coded on the reverse contig strand.

Table 2.3. Summary of protein identifications

				DEP* origins			
	Identified	Quantified	DEP	Varroa	Virus	Bee	New
Development	3,102	2,626	1,433	1,148	2	114	169
Sex	3,000	2,260	101	86	0	1	14

*DEP: differentially expressed protein

2.5.2 **Carbohydrate metabolism enzymes are strongly developmentally regulated**

To gain a better understanding of which cellular processes are regulated by proteins that were differentially expressed through development, we performed an enrichment analysis by gene score resampling (GSR) and found, not surprisingly, that lipid localization and lipid transport were among the most significantly enriched (Table 2.4), driven largely by vitellogenin expression. Many processes involved in aerobic respiration were also significantly enriched, including GO terms linked to glycolysis (GO:0006090, GO:0006096) and the citric acid cycle (GO:006099, GO:0072350). To investigate these metabolic processes further, we analyzed how the abundances of core glycolysis and citric acid cycle enzymes varied with development (Figure 2.12 A). Most enzymes (16/20) were significantly differentially expressed and only two (phosphoglyceromutase and succinyl CoA synthetase) were not quantifiable. Several enzymes appear to have multiple isoforms, based on BLAST search results, some of which are not co-expressed (*e.g.* for hexokinase, α-ketoglutarate dehydrogenase, aconitase, isocitrate dehydrogenase and malate dehydrogenase). Overall, the foundress mite has the highest levels of most enzymes, and when this is not the case it is largely due to age-specific isoform expression.

Table 2.4. GO terms significantly enriched in developmental stages

Category	Name	ID	# Genes	p value*
Glycolysis & TCA	aerobic respiration	GO:0009060	15	0.0777
	carbohydrate metabolic process	GO:0005975	62	0.0556
	dicarboxylic acid metabolic process	GO:0043648	10	0.0551
	tricarboxylic acid cycle	GO:0006099	13	0.0509
	glycolytic process	GO:0006096	10	0.0526
	cellular respiration	GO:0045333	22	0.0531
	monocarboxylic acid metabolic process	GO:0032787	36	0.0358
	energy derivation by oxidation of organic compounds	GO:0015980	24	0.0343
	tricarboxylic acid metabolic process	GO:0072350	14	0.0288
	pyruvate metabolic process	GO:0006090	16	0.0179
	generation of precursor metabolites & energy	GO:0006091	36	4.5E-10
	ADP metabolic process	GO:0046031	11	0.0544
	ATP metabolic process	GO:0046034	32	7.5E-11
	ribonucleoside triphosphate metabolic process	GO:0009199	35	1.1E-10
	purine nucleoside triphosphate metabolic process	GO:0009144	34	1.5E-10
	nucleotide phosphorylation	GO:0046939	16	0.0056
	nucleoside triphosphate metabolic process	GO:0009141	37	0.0064
	purine nucleoside monophosphate metabolic process	GO:0009126	45	0.0149
	purine nucleotide metabolic process	GO:0006163	49	0.0244
	ribose phosphate metabolic process	GO:0019693	57	0.0261
	purine-containing compound metabolic process	GO:0072521	51	0.0269
	nucleoside monophosphate metabolic process	GO:0009123	48	0.0280
	ribonucleoside monophosphate metabolic process	GO:0009161	47	0.0310
	ribonucleotide metabolic process	GO:0009259	51	0.0348
	nucleoside diphosphate phosphorylation	GO:0006165	12	0.0354
Amino acid metabolism	aromatic amino acid family metabolic process	GO:0009072	7	0.0627
	cellular amino acid metabolic process	GO:0006520	63	0.0512
Lipid movement	lipid localization	GO:0010876	9	9.0E-11
	lipid transport	GO:0006869	9	2.2E-10
Electron transport chain	electron transport chain	GO:0022900	12	0.0694
	ATP biosynthetic process	GO:0006754	10	0.0973
Chemical homeostasis	chemical homeostasis	GO:0048878	14	0.0504
	cellular chemical homeostasis	GO:0055082	8	0.0720
Cation transport	cation transmembrane transport	GO:0098655	35	0.0981
	cation transport	GO:0006812	42	0.0728
Other	protein deubiquitination	GO:0016579	8	0.0999
	intra-Golgi vesicle-mediated transport	GO:0006891	5	0.0587

*Benjamini-Hochberg corrected enrichment p value

2.5.3 **Changes in cuticle-associated proteins throughout development**

Many proteins related to cuticle formation did not map to GO terms, despite having significant BLAST hits to chitin-like proteins. To analyze how cuticle formation may be developmentally regulated, we manually retrieved all proteins with significant BLAST hits to chitinases, structural chitin and chitin binding proteins. Indeed, we observed stark differences in the types of chitinases and structural chitin that are utilized (Figure 2.12 B**).** Young mites displayed a markedly different structural chitin profile than adult sons and daughters, which was different still compared to the armoured foundress.

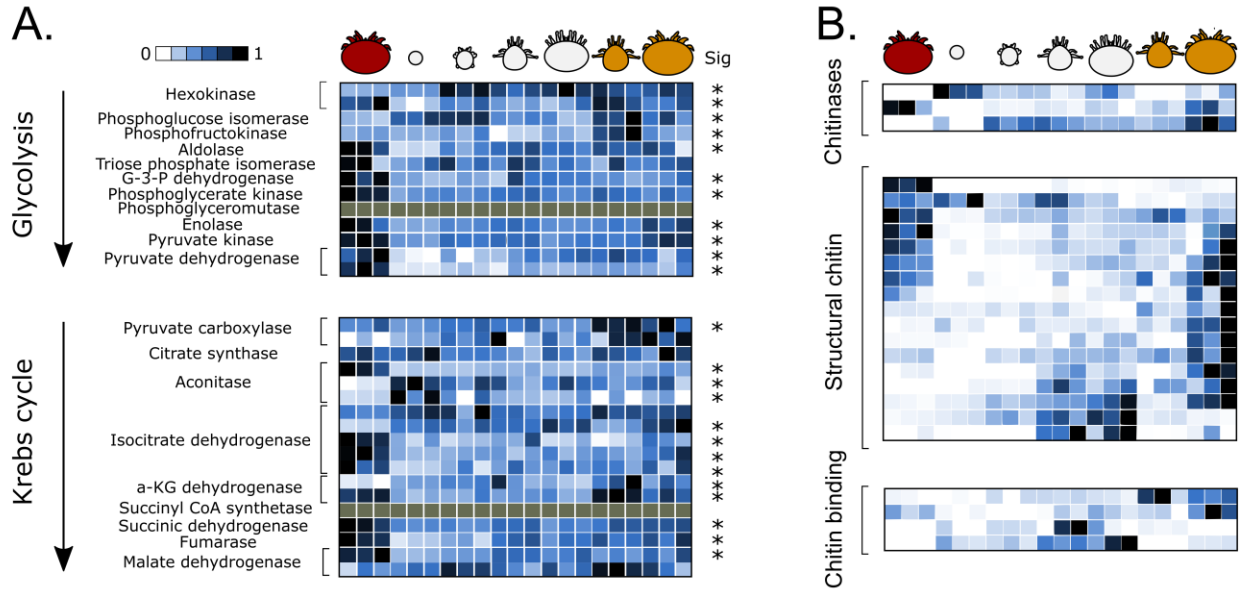


Figure 2.12. Analysis of carbohydrate metabolism enzymes and cuticle proteins.

A. Relative expression of enzymes involved in carbohydrate metabolism. Bracketed rows indicate isoforms of enzymes catalyzing the same reaction (based on shared enzyme codes and having the enzyme in question as the best BLAST hit). Grey tiles indicate the protein was not observed. Rows indicated with an asterisk are significantly differentially expressed across developmental stages. G-3-P: glyceraldehyde 3-phosphate; a-KG: α-ketoglutarate. B. Relative expression of proteins related to chitin formation. Only the significantly differentially expressed proteins are shown.

2.5.4 **Sexually regulated proteins**

Varroa follows the system of haplodiploid sex determination (*i.e.* females are diploid, males are haploid), but other than that very little is known about the mechanisms that contribute to sexual differentiation. To investigate this, we compared the proteins expressed in female (n = 9) and male (n = 6) mites and found 101 starkly differentially regulated proteins, providing a starting point on which to further investigate possible differentiation mechanisms (Figure 2.13 A). A disproportionately large fraction (over 80%) of the differentially regulated proteins were upregulated in the males. Investigating the 10 most significant proteins further, we found that only three had appreciable homology to sequences with known functions (Figure 2.13 B) – uridine phosphorylase, histone lysine N-methyltransferase and heat-shock protein (HSP)83 – while the others either had no significant sequence similarities or the significant matches have not been functionally annotated. Despite this, functional enrichment analysis revealed that GO terms relating to chromatin remodeling, positive regulation of transcription as well as various metabolic processes were significantly enriched (Table 2.5). Intrigued by the prominent profile of HSP83, we further analyzed how the other HSPs are sexually regulated (Figure 2.13 C). We found that there is a core group of three HSPs that are specific to the foundress, and another group of three HSPs are male-specific.

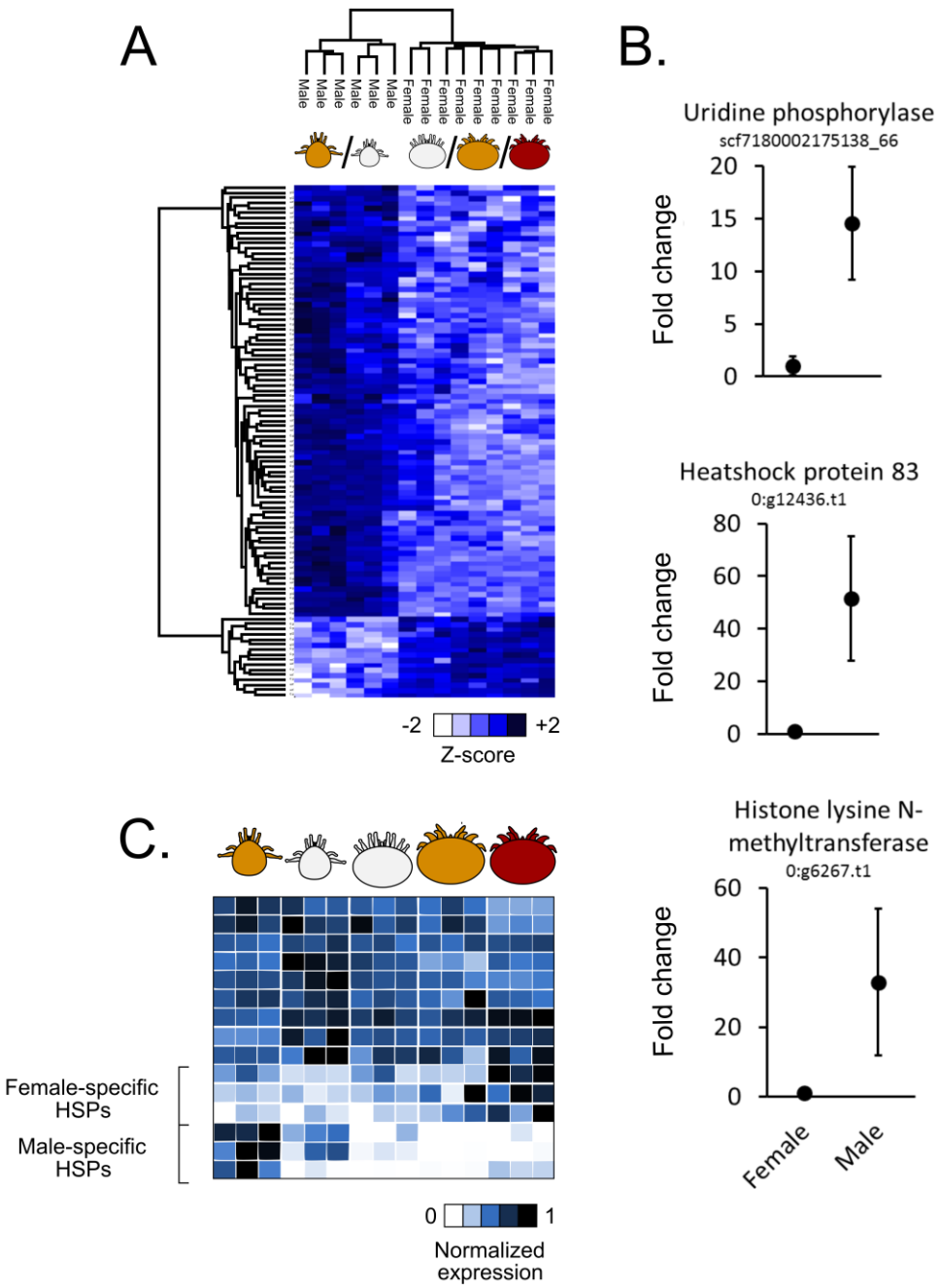


Figure 2.13. Sexually regulated proteins in Varroa.

A. Heatmap showing differentially expressed proteins in male (n = 6) and female (n = 9) mites (Benjamini Hochberg-corrected FDR = 5%). Hierarchical clustering was performed using average Euclidian distance (300 clusters, maximum 10 iterations). B. The proteins with known functions among the top 10 differentially expressed. Fold change is normalized to the average expression in females. Error bars are standard deviation. C. Relative expression of heat-shock proteins (HSPs). Each row represents one HSP. Only significantly differentially expressed HSPs are shown.

Table 2.5. GO terms significantly enriched in sexually regulated proteins

Description	Name	ID	# Genes	p value*
Chromatin remodeling	DNA packaging	GO:0006323	7	0.0970
	DNA conformation change	GO:0071103	11	0.0698
	chromatin assembly or disassembly	GO:0006333	6	0.0833
Transcription	positive regulation of gene expression	GO:0010628	6	0.0882
	positive regulation of transcription, DNA-templated	GO:0045893	5	0.0945
Biosynthesis	positive regulation of biosynthetic process	GO:0009891	6	0.0950
	aromatic compound biosynthetic process	GO:0019438	70	0.0953
	organophosphate biosynthetic process	GO:0090407	43	0.0967
	nucleotide biosynthetic process	GO:0009165	37	0.0919
	glutamine family amino acid biosynthetic process	GO:0009084	6	0.0750
	organic acid biosynthetic process	GO:0016053	24	0.0798
Metabolism/catabolism	cellular amino acid metabolic process	GO:0006520	64	0.0748
	alcohol metabolic process	GO:0006066	5	0.0838
	cellular nitrogen compound catabolic process	GO:0044270	13	0.0882
	organic cyclic compound catabolic process	GO:1901361	16	0.0882
	glycosyl compound metabolic process	GO:1901657	21	0.0882
	nucleoside metabolic process	GO:0009116	20	0.0907
	heterocycle catabolic process	GO:0046700	14	0.0937
	nucleobase-containing compound catabolic process	GO:0034655	9	0.0937
Other	peptidyl-amino acid modification	GO:0018193	34	4.4E-10
	response to organic substance	GO:0010033	7	0.0682
	response to oxygen-containing compound	GO:1901700	5	0.0762
	protein folding	GO:0006457	48	0.0882

* Benjamini Hochberg corrected enrichment p value

2.6 Discussion

The work presented here provides a foundation to begin to unravel the fundamentals of *Varroa* biology, including developmental transitions, sexual differentiation, and host-virus interactions. Both the newly generated data (in the case of *Varroa*) and re-analysis of existing data (in the case of honey bees) will also assist with improving the genome annotation in future iterations.

2.6.1 **Proteogenomics efforts**

Genome sequencing has become relatively easy, but accurately annotating the genome is an arduous and imperfect process. The most common model organisms (*e.g. M. musculus*, *D. melanogaster*, *C. elegans*, *etc.*) have benefitted from decades of genetic research which has refined their genome annotations over time, resulting in highly reliable and accurate gene sets on which most tools for analyzing global gene and protein expression rely. Our data clearly show that the new *Varroa* gene annotation is far better than the provisional draft (Figure 2.2), but our proteogenomics initiative, which still identified further unique, unannotated candidate protein-coding sequences, suggests that there is still room for improvement. Likewise, our proteogenomics investigation in honey bees suggests the same, despite it being a more extensively researched organism than *Varroa*.

Our intention from the outset was to produce an inclusive list of potentially expressed regions, then narrow that list down from there (rather than apply strict criteria in the beginning). Therefore, when we predicted candidate ORFs from the genome and transcriptome sequences, we only required ORFs to be a minimum of ninety nucleotides, which may mean that the smaller coding regions identified here could be additional exons from known genes, rather than new, fully independent proteins. This is difficult to ascertain from the current version of genome sequences for these species, however, because there are still too many unscaffolded contigs, so further work will need to be done to confirm the exact location of splice junctions and which isoforms are expressed. In fact, the current honey bee and *Varroa* annotations do not include any splice variants, so any information such as this would be a significant asset. A limitation of

proteogenomics by mass spectrometry is that the methods of sample preparation (protease digestion) obscures the physical linkage between exons, so this technology alone is not sufficient to fully annotate splice isoforms. For the same reasons, this technology cannot distinguish between merged genes (separate genes annotated as one); however, it could be used in conjunction with long read transcriptome sequencing to further refine the genome annotation in honey bees and other organisms.

The annotation process is not only influenced by the genome itself (chemical and physical properties, completeness, *etc*), but also by the quality of guiding transcript assemblies and a number of human-determined parameters (*e.g.* the annotation software employed, hard or soft repeat masking, splice site awareness, *etc.*), and availability of prior gene models[188,189]. Furthermore, some parameters may need to be altered on a species-by-species basis, but there is no inherent pathway for finding the optimal settings. Proteomics and RNA-seq data could serve as tools to not only confirm expression of predicted genes, but also to help define these parameters in the first place since the resulting protein and gene identifications are sensitive to database accuracy. We feel that both species investigated here could benefit from a concerted, community-wide re-annotation effort using all available expression data to more accurately predict gene models.

2.6.2 **Developmentally regulated proteins in *Varroa***

Our analysis of developmentally-regulated proteins in *Varroa* revealed some intriguing trends regarding the energetic demands throughout the mite's development. The foundress had consistently high abundances of enzymes that participate in glycolysis and the citric

acid cycle, which may be required to meet the energetic demands of producing and laying eggs. We speculate that many of the differences in metabolic processes are also driven by the unique energetic requirements of metamorphosis, when energetically expensive morphological rearrangements must occur while the mite does not eat.

During maturation, protonymph and deutonymph mites transition from having a soft, translucent cuticle to acquiring a harder and more durable exoskeleton. The phoretic and foundress mites have rigid armour to protect against injury by grooming honey bees and other environmental hazards. To investigate the possible mechanisms behind these transitions, we compared the expression profiles of significantly differentially expressed proteins that are related to cuticle development (chitin structural proteins, chitinases and chitin binding proteins). The egg contains large amounts of one chitinase and one chitin structural protein, which could be related to the breakdown of the egg case or the developing mite larva as it becomes a protonymph. Deutonymphs display a specific profile of highly abundant structural proteins and chitin binding proteins, and from this point on there is a clear separation between male and female expression profiles. While foundress mites are armoured, the male mite, however, appears not to invest energy in forming a tough exoskeleton like the female does – based on their general lack of chitin-associated protein expression – which is consistent with the lack of environmental exposure during the male life cycle. This represents just one potential target that we could disrupt in order to control *Varroa* populations. If we can target the structural proteins responsible for

exoskeletal formation, either by RNAi or genetic engineering, we could make the mites more vulnerable to honey bee grooming or possibly to existing acaricides.

2.6.3 **Sexually regulated proteins**

In our analysis of sexually regulated proteins, we found that chromatin remodelling and transcription activation were significantly enriched processes. Chromatin remodeling could be required to de-condense chromosomal regions which are highly expressed in males or females, and vice versa. Indeed, histone lysine N-methyltransferase was one of the most significant differentially expressed proteins, with approximately 30-fold higher levels in males compared to females (Figure 2.13 B) and peptidyl-amino acid modification was the most significantly enriched biological process (Table 2.5). This kind of "on-off" regulation could thus be very important for sex determination. We also found that HSP83, which is critically important for spermatogenesis in *Drosophila*[190], displayed the greatest fold change (~50-fold) out of those with known functions. Broadening our analysis to all identified HSPs, we found that there is a core group of HSPs that are specific to the foundress and another group that is specific to males (Figure 2.13 C), suggesting that these HSPs are involved in regulating the transcription of sex-specific genes.

2.6.4 **Conclusion**

The work we present here represents a first-of-its-kind, high-resolution analysis of the *Varroa* proteome. With some 1,433 proteins that are differentially expressed, this data

provides a first glimpse into the changes that take place during *Varroa* development. In addition, the strongly sexually regulated proteins provide clues for discovering the mechanisms behind sex determination and general dimorphism. We hope that the interactive web tool will maximize the utility of this information for the research community and will help generate further hypotheses for future experiments on this major honey bee pest.

Furthermore, through our honey bee and *Varroa* proteogenomic analyses, we found evidence for many undescribed candidate protein-coding regions in both species. This includes 1,426 candidate ORFs in honey bees which were supported by two or more peptides, with thirty percent of all new proposed protein-coding regions supported by transcript evidence from a single RNA-seq data set. Several sequences were validated individually. For *Varroa,* we identified 301 new proposed protein-coding regions using only our in-house generated mass spectrometry data set, many of which were actually differentially expressed between life stages. Although we cannot say for sure which of these are new independent coding regions and which are exons of known genes, this is a step towards refining the completeness of the honey bee and *Varroa* genome annotations. We urge researchers to integrate these proteogenomics peptide data into future annotation refinement efforts for both *Varroa* and honey bees.

2.7 Methods

2.7.1 *Varroa* and honey bee data access

The *Varroa* proteomics data is available on ProteomeXchange (www.proteomexchange.org; accession: PXD006072). The honey bee proteomics data is available at PXD002275 and the Honey Bee Peptide Atlas (www.peptideatlas.org).

2.7.2 *Varroa* samples

2.7.2.1 *Varroa* sample collection

Varroa mite families were collected from a single honey bee colony in the fall of 2016 in Vancouver, Canada. In a large-scale population genomics study, the authors found that the genetic variation of *Varroa* within colonies accounted for by far the largest fraction of genetic variation compared to between colonies and between apiaries[191]; therefore, sampling mites from a single colony was sufficient. Eggs, foundresses, adult daughters and adult sons were transferred directly to microfuge tubes using a soft paintbrush, whereas protonymphs and deutonymphs were transferred to a petri dish and sorted under a dissecting microscope according to the identification guides available at http://idtools.org/id/mites/beemites and http://extension.msstate.edu/publications (publication number: P2826) *via* the University of Michigan and the Mississippi State University, respectively. Approximately 50 individuals were pooled for each replicate (7 developmental stages, n = 3 for each stage). All samples were immediately frozen at -72 ºC until protein extraction.

2.7.2.2 *Varroa* protein preparation

Protein was extracted by homogenizing each mite stage with ceramic beads as previously described[192]. Clarified lysate was precipitated overnight with 4 volumes of 100% ice cold acetone and the pellet was washed twice with ice cold 80% acetone. After allowing residual acetone to evaporate (~15 min) the protein pellet was solubilized in urea buffer (6M urea, 2M thiourea in 10 mM HEPES, pH 8) and ~30 µg (determined via the Bradford Assay) was reduced, alkylated and digested with Lys-C, then trypsin as previously described[193]. Peptides were acidified (1 volume 1% TFA), desalted on a high capacity C18 STAGE tip[194], solubilized in Buffer A (0.1% formic acid) and quantified in technical triplicate using a peptide fluorometric assay (Pierce; cat: 23290).

2.7.2.3 *Varroa* data acquisition

Two µg of peptides per sample were analyzed on an EasynLC-1000 chromatography system (Thermo) coupled to a Bruker Impact II Q-TOF mass spectrometer. The LC C18 columns included a fritted trap column and pulled-tip, 50 cm analytical column produced and packed in-house[151,195]. Peptides were separated using a 165 min linear gradient of increasing Buffer B as specified in the LCParms.txt file embedded within the Bruker data folders (available at www.proteomexchange.org, accession: PXD006072). Buffers A and B were 0.1% formic acid and 0.1% formic acid, 80% acetonitrile, respectively. The instrument was set to the same parameters as described in our previous publication under "Analysis of PTMs"[151], except the scanned mass range was 200-2,000 m/z, the top 20

precursors were fragmented at a 5 Hz spectral rate and the lower precursor intensity threshold was 300 counts.

2.7.2.4 *Varroa* proteogenomics

For the proteogenomics analysis, the *Varroa* spectra were searched against a six-frame translation of the publicly available *Varroa* genome sequence (PRJNA33465) using MaxQuant (v.1.5.3.30) to identify new protein-coding regions (minimum ORF length was set to 100 amino acids). All viruses known to infect honey bees and *Varroa* were also included in the database. Honey bee proteins were not included after a follow-up sequence similarity analysis indicated that only 5 of the proteins identified in this search matched to bees. MaxQuant search settings included: trypsin cleavage specificity, 2 allowed missed cleavages, fixed carbamidomethyl modification, variable oxidated methionine and N-terminal acetylation, 0.07 Da precursor mass tolerance, 35 ppm fragment mass tolerance, and 1% protein and peptide FDR calculated based on reverse hits. The peptide (scores, modifications, precursor mass and m/z) and protein (protein groups, accessions, number of assigned peptides, unique peptides and % coverage) identification information contained within the main MaxQuant output files (summary.txt, peptides.txt, proteinGroups.txt, parameters.txt) and the protein database (165,951 entries) are available at PXD006072. Annotated spectra are available through MS-viewer (search key: wuh30b9smr).

Peptides identified in the six-frame translation search but which were not present in the canonical protein database were used as anchors to retrieve the corresponding open

reading frames from the genome using a simple Perl script. This yielded 524 candidate new protein-coding sequences. Of these, 301 were flanked by two or more peptides spanning at least 50 amino acid residues. We used a two-way ANOVA (factors: amino acid and new/known sequence origin) to compare amino acid composition between this set of 301 candidate protein-coding sequences and 902 sequences bounded by known peptides that were identified in the same six-frame translation search. We used these 902 sequences, which were also generated by the MaxQuant six-frame translation algorithm, because protein-coding sequences generated by more sophisticated algorithms (as with the canonical Varroa annotation) could generate different sequence properties simply due to the algorithm being different. These 902 sequences, however, were both a product of the six-frame translation and part of the canonical protein database. Next, we used the same approach to compare nucleotide positions within codons (factors: nucleotide position and sequence origin). We also compared the AT frequency of the new coding regions, known coding regions and genome sequences that were broken into 1 kb segments *in silico* (n = 384,129) using a one-way ANOVA (3 levels) with a Tukey HSD post-hoc test.

To survey these proteins for orthology with other species and to retrieve GO terms, we performed Blast2GO (v4.0) using default parameters. We reasoned that these sequences might have been missed in the *Varroa* annotation effort if they only share sequence similarity to evolutionarily distant species; therefore, we queried them against the non-redundant protein collection with no taxonomic restrictions. Five sequences showed significant homology to honey bee sequences and were removed from the list of new protein sequences, leaving 519 in total.

2.7.2.5 Label-free protein quantification

We searched the mass spectrometry data using the same parameters as above, except label-free quantitation (LFQ) was enabled (with min ratio count = 1) and a composite protein database was used which included all proteins in the most recent *Varroa* gene annotation (the final protein database is included at the ProteomeXchange accession), the 519 candidate protein sequences identified above, all viral sequences known to infect honey bees or *Varroa* and all proteins contained within the honey bee OGSv3.2 annotation. Since honey bee biological material is *Varroa'*s sole food source, we expected to find a substantial number of honey bee proteins within our samples. The final database totaled 32,110 entries and is available at PXD006072, along with the MaxQuant peptide and protein identification information. Honey bee proteins include an "Amel" tag in the accession, new protein-coding regions from the six-frame translation include a "True" or "False" tag in the accession (indicating the DNA template strand relative to the indicated contig), virus sequences are represented by a single gi number or Uniprot identifier and all other sequences (excluding contaminants and reverse hits) belong to *Varroa*. Annotated spectra are available at MS-viewer (search key: msmx6z444s).

2.7.2.6 Functional enrichment analysis

We performed functional enrichment analysis on two sets of proteins: 1) *Varroa* proteins that were differentially expressed through development and 2) *Varroa* proteins that were differentially expressed between sexes. For all protein sets, we retrieved GO terms using Blast2GO (v4.0) with default parameters, first searching against all arthropods, then

sequences with missing GO terms were searched again against the entire non-redundant protein collection. GO terms were exported after running the GO-Slim function. We then performed a gene score resampling (GSR) analysis with ErmineJ v3.0.2[196], using log-transformed q values (from the previous differential expression analysis) for "protein score." We considered a GO term significantly enriched if the Benjamini Hochberg-corrected GSR p-value was less than 0.10.

2.7.2.7 Building the *Varroa* protein atlas

The web-based interactive *Varroa* protein atlas was built using the framework previously described for the honey bee protein atlas[125].

2.7.3 Honey bee samples

2.7.3.1 Honey bee sample collection

With the exception of the sequenced worker honey bee, all samples were collected from colonies at the UBC apiary (Vancouver, British Columbia, Canada) between April and August 2014. The sequenced honey bee was collected during the summer of 2010 from the York University Apiary (Toronto, Canada).

2.7.3.2 Honey bee, yeast, human, mouse, fly, and worm protein preparation

Samples comprised of *Apis mellifera* (5th instar worker larvae), *Saccharomyces cerevisiae*, *Homo sapiens* (HeLa cells), *Mus musculus* (liver), *Drosophila melanogaster* (adult females), and *Caenorhabditis elegans* (mixed life stages; N = 3 each). Protein was extracted in 50 mM ammonium bicarbonate (1% sodium deoxycholate) with the aid of ceramic beads for coarse

tissues. Lysate was boiled at 99˚C for 5 minutes, debris was pelleted, then protein concentrations were determined using a bicinchoninic acid (BCA) assay (Pierce) and confirmed by SDS-PAGE. Protein was reduced (1.0 µg DTT/50 µg protein, 37˚C, 30 min), alkylated (5 µg iodoacetamide/50 µg protein, 37˚C, 20 min) and digested overnight with porcine modified trypsin (Promega; 1 µg trypsin/50µg protein, 37˚C, 16 h). Samples were acidified with 1% TFA solution until pH < 2.0 and the precipitated deoxycholate was removed by centrifuging through a 0.6 µm filter (Sartorius Stedim Biotech). Peptides were then fractionated (x5) by strong cation exchange as described previously[197].

2.7.3.3 Data acquisition

Digested peptides were analyzed by LC-MSMS using a nanoflow HPLC (Thermo easy-nLC1000) coupled to a Q-Exactive mass spectrometer (Thermo). For each sample, approximately half of each SCX fraction was injected into the LC and loaded onto an in-house packed fused-silica (5 µm Aqua C18 particles (Phenomenex)) fritted trap column (2 cm, 100 µm I.D., 360 µm O.D., 5 µL/min flow rate, Buffer A = 0.5% acetic acid), then resolved on a reverse phase 75 µm inner diameter fused silica, in-house packed 30 cm analytical column (ReproSil C18, 3 µm particle size (Dr. Maisch)) using a 75 min linear gradient run at 250 µl/min from 5 % to 35% Buffer B (acetonitrile, 0.5% acetic acid), followed by a 15 min wash at 95% Buffer B. Instrument acquisition parameters included a 1% underfill ratio, 70,000 precursor mass resolution, 17,500 fragment mass resolution, normalized collision energy of 28%, +1 and unassigned charges were excluded, “exclude isotopes” was turned on, intensity dependent MSMS at 1.7e5 intensity threshold, and the instrument was set to scan from 300 to 2000 m/z with a 30 s dynamic exclusion time.

2.7.3.4 Species comparison data processing

Data were searched using MaxQuant (v1.5.2.8). Parameters included: carbamidomethylated cysteine (fixed), methionine oxidation (variable) and protein N-terminal acetylation (variable) modifications; trypsin specificity unless otherwise stated; maximum two missed cleavages; 10 ppm precursor mass tolerance; 0.05 Da fragment mass tolerance; 1% FDR; +1 to +7 charge states; match between runs was enabled, and; common contaminants were included. The honey bee protein database OGSv3.2 (15,314 sequence entries), which is the most recent OGS, was downloaded from BeeBase (www.beebase.org). Non-bee databases were downloaded from Uniprot (including fragment sequences; *H. sapiens:* 135,735 entries; *M. musculus:* 74,182 entries; *S. cerevisiae:* 6,629; *C. elegans:* 26,448, *D. melanogaster:* 20,049). Student's t-tests were performed on the average protein identifications as well as proteome coverage (*i.e.* fraction of protein-coding genes observed) using the Microsoft Excel 2013 data analysis package.

2.7.3.5 Honey bee proteogenomics

To provide support for new coding regions, we used the entire honey bee Peptide Atlas dataset (~1,500 raw files, publicly available at www.peptideatlas.org, accessed 2015/02) to search three different databases. This included a six-frame genome (build: Amel_4.5) translation database (yielding 216 million residues) produced using MaxQuant v1.4.1.2 using a minimum ORF size of 30 amino acids. The intention for this search was to identify new candidate protein-coding regions that can be included in training gene sets for future iterations of annotation refinement. The data was also searched against a three-frame transcriptome translation database (NCBI ref-Seq, yielding 166 million residues) because we predicted that we may get more hits owing to the

~50% reduction in search space compared to the genome translation database. Finally, the Peptide Atlas data set was also searched against a database including all unique protein sequences contained in OGSv1.0 and OGSv3.2 to determine if any of the ~6,000 genes that were lost during the database upgrade should be reconsidered.

In order to provide further support for the expression of candidate coding regions identified in these searches, we searched for amino acid sequence similarity to other organisms. We used BLAST+ v2.2.30 to align all ORFs covered by at least one peptide (relaxed) or two peptides (strict) at 1% FDR to sequences in the complete non-redundant protein database (all species). All parameters were left as default, and the top ten alignments were outputted.

Chapter 3: Identifying candidate hygienic behaviour-inducing odorants[3]

3.1 Introduction and rationale

Honey bees face many challenges, but disease is perhaps the most significant[198] and hygienic behaviour is an important method of disease control[48-50,52,87]. The most well-established method of selecting for hygienic bees is the freeze-killed brood (FKB) assay, in which patches of brood are frozen with liquid nitrogen, killing them but otherwise leaving them undamaged, and returned to the hive to be evaluated after 24 h. The hygienic score is defined as the fraction of dead pupae that have been detected and removed[47,49]. Colonies that perform well in this test also have improved outcomes when challenged with real diseases (*i.e. Varroa*, American foulbrood, and chalkbrood), allowing the FKB assay to be an effective tool for selective breeding[47].

There is a large body of evidence suggesting that hygienic bees identify dead and diseased brood through olfactory cues[39,91-94,199,200], and that they are more sensitive to and better at discriminating between them[91,200]. The antennae, bees' main olfactory appendages[201], have been shown to play a pivotal role in hygienic behaviour with multiple independent research groups identifying significantly differentially expressed antennal genes in hygienic *versus* non-hygienic bees, as well as strong antennal biomarkers for selective breeding[37,39,115,199,202]. Furthermore, odorant binding proteins (OBPs) aid odorant detection and are consistently upregulated in hygienic bees' antennae. However, relatively little is known about precisely what odorants the bees are detecting. One study investigated the volatile odorants emitted from chalkbrood-infected

[3] Content in this chapter has been published: McAfee, A. *et al*. Odorant cues linked to social immunity induce lateralized antenna stimulation in honey bees (*Apis mellifera* L.). *Sci. Rep*. 7:46171. Copyright (2017) Authors. Permission not required for reprinting. This work is licensed under a Creative Commons Attribution 4.0 International License.

larvae[94] and several focused on possible cues from *Varroa*-infested brood[95,96,203-205], but none investigated how they compare to FKB (the main selective test for hygienic behaviour) and very few confirmed the biological activity of the compounds[94,95]. This seems to be a missed opportunity since studying hygienic behaviour mechanisms against FKB should be a good model system, enabling standardized freeze treatments and avoiding experimental infections. Despite the interest in finding *Varroa*-associated hygienic behaviour inducers, only one specific molecule has been suggested and investigated with bioassays (*Z*-(7)-pentadecene), but there is not strong evidence for its differential emission from infested brood. Furthermore, how infested brood odorant profiles change with respect to pupal development (and associated growth of the *Varroa* mite family) is yet unknown. Non-volatile cues have not yet been investigated behaviourally in honey bees. Despite the prediction that non-volatile compounds should poorly penetrate the wax brood cap, they could still stimulate hygienic behaviour if either the bees are very sensitive to it, or if they are attracted to investigate the cell for some other primary reason, at which time they may come in contact with the non-volatile compound.

3.2 Research goals

In this study, our main goals are to identify candidate hygienic behaviour-inducing odorant signals emitted from two distinct 'disease' states – *Varroa*-infested brood and FKB – relative to healthy brood. We first aimed to do this using gas chromatography-coupled mass spectrometry (GC-MS) to compare odorant profiles of FKB versus age-matched healthy brood across pupal developmental stages within one colony (Figure 3.1 A). Our second aim is to use the same technique to conduct a cross-colony comparison of the FKB versus healthy brood for a

developmental stage often utilized for the FKB assay (Figure 3.1 B). Our third aim is to make similar comparisons between *Varroa*-infested brood and healthy brood, also for multiple developmental stages (Figure 3.1 C). Finally, our fourth aim is to functionally validate the biological activity of candidate hygienic behaviour-inducing compounds. Our approach is to use electroantennography to quantify the strength by which they stimulate antennae of hygienic and non-hygienic bees. Overall, the data we produce is essentially a broad screen for candidate hygienic behaviour-inducing compounds, which will require further behavioural validation. These data will be an important foundation on which to build our understanding the molecular mechanism behind this complex behaviour.

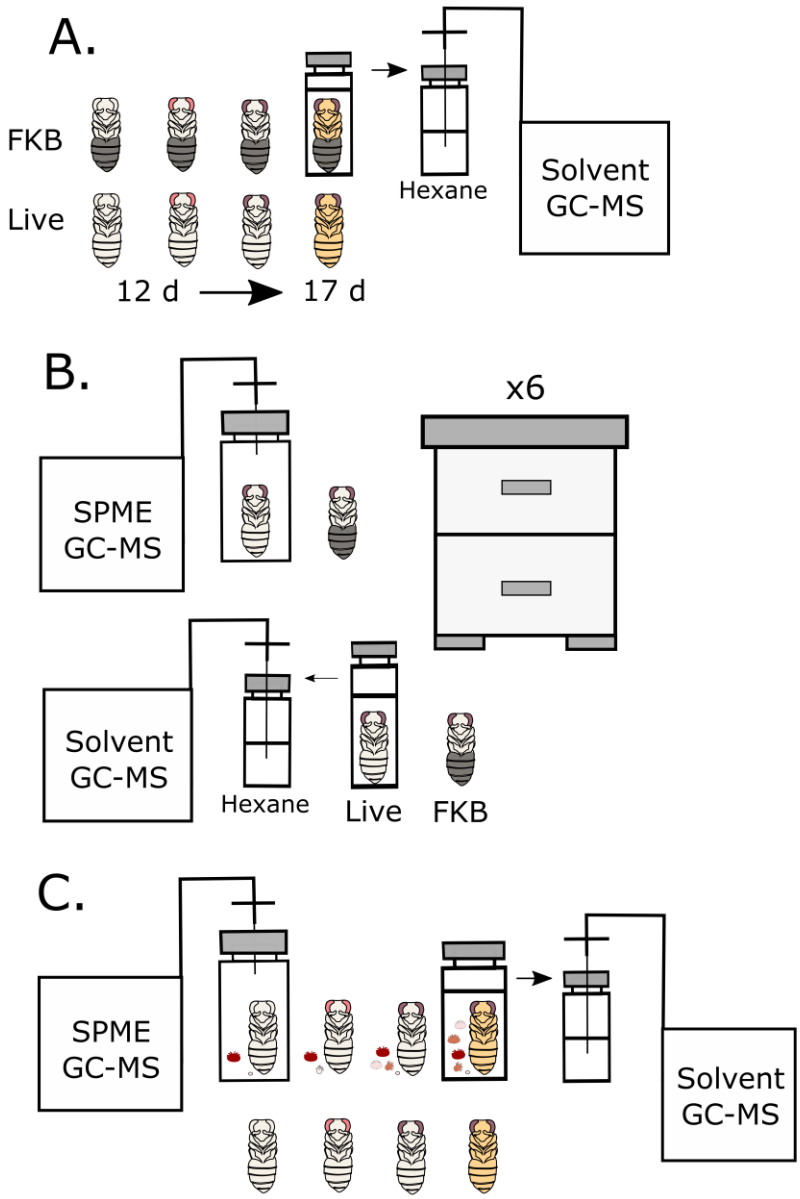


Figure 3.1. Experimental design schematic for studies investigating freeze-killed brood and *Varroa*-infested brood odorant profiles.

A. Cuticular hydrocarbon analysis. N = 3 individuals for each developmental stage (white-eyed, pink-eyed, purple-eyed white body, purple-eyed tan body), sourced from one colony. FKB: Freeze-killed brood; GC-MS: gas chromatography mass spectrometry. B. Cross-colony comparison of headspace volatiles and cuticular hydrocarbons. N = 3 for each colony. SPME: solid phase microextraction. C. *Varroa*-infested brood headspace volatiles and cuticular hydrocarbons. Mites and their families were included in each sample. N = 3 for each developmental stage, sourced from a single mite-infested colony.

3.3 Emission of odorants from freeze-killed brood

3.3.1 Candidate FKB odorant cues

The FKB assay is thought to work best on young (white-eyed to red-eyed) pupae, but the reason behind this is unknown. It is possible that younger FKB could emit a different odorant profile than older FKB. To investigate this, we used GC-MS to roughly compare the cuticle molecular profiles of 12 to 17 d old pupae at 1 to 2 d intervals (Figure 3.1 A). We did not use internal standards or produce a calibration curve to aid in quantification; rather, we compared compound peak areas directly. While this method could lead to an underestimation of abundance if a compound approaches saturation, it offers a simple, fast screen to identify the most prominent differences that might be biologically significant. We found that indeed there were strong differences between dead and live brood (three-factor ANOVA; $p < 0.000001$; $F = 597$), which interacted significantly with developmental stage ($p = 0.0000024$; $F = 9.72$) and compound identity ($p < 0.000001$; $F = 10.7$). Young (12 to 15 d old) FKB tended to have more differentially emitted compounds compared to old (16 to 18 d) FKB (Figure. 3.2), which could explain why performing the FKB assay on patches of younger pupae appears to be more sensitive. While most of these compounds were age-specific, one compound, oleic acid, was consistently different across all ages. The identity of this compound was confirmed against a commercial standard (Table 3.1).

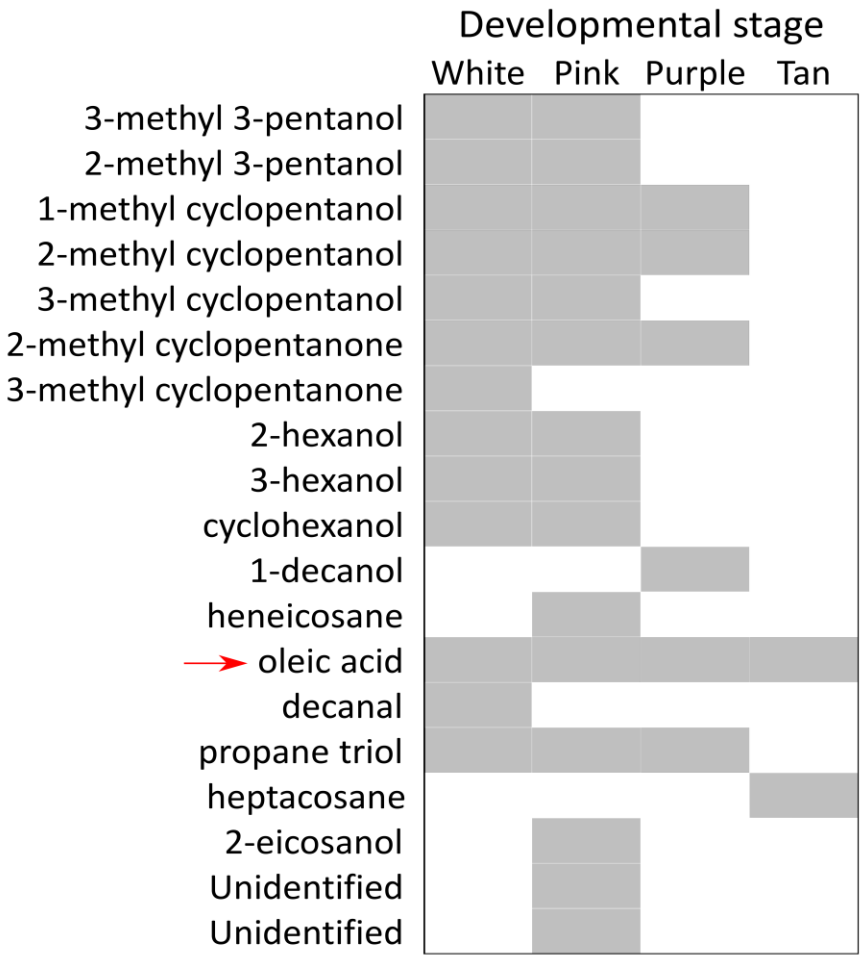


Figure 3.2. FKB-specific odour profiles vary across developmental stages.

Cuticle compounds from live and freeze-killed white-eyed (12-13 d), pink-eyed (14-15 d), purple-eyed white body (16 d) and purple-eyed tan body (17-18 d) were analyzed using gas chromatography mass spectrometry (N = 3 for each age). Shaded boxes indicate compounds which had statistically significantly different abundances in FKB compared to age-matched live brood. Compounds were identified by comparing GC-MS spectra against a compound library. Only the identity of oleic acid (red arrow) was confirmed against a synthetic standard.

Table 3.1. Differentially emitted compound identifications

Compound[1]	Base peak (m/z)	Score[2]	Sample RT (min)	Standard RT (min)	Identification accuracy[3]	P value
Isopropanol	44.99	90.6	3.0	3.0	High	2.2E-23
2-pentanone	43.01	91.4	3.8	3.9	High	3.4E-23
E-β-ocimene	93.00	94.2	12.6	12.7	High	2.3E-33
Oleic acid	55.07	72.6	25.0	25.1	High	1.8E-70
Compound 1 (α-thujene)	93.00	82.3	12.1	5.5	Low	9.5E-10
Compound 2 (α-pinene)	93.00	94.4	12.4	5.6	Low	2.5E-12
Compound 3 (2,3-butanediol)	42.98	79.2	13.3	22.4	Low	1.2E-13
Compound 4 (2-methyl tetradecane)	57.09	83.5	14.7	N/A	Low	1.6E-45

[1] Bracketed compound names represent the proposed Mass Hunter matches
[2] Mass Hunter Qualitative Analysis v.B.06.00
[3] Based on comparison to synthetic standards

We hypothesized that the compounds most likely to be hygienic behaviour-inducers should also be consistently differentially emitted from dead brood across different colonies. We compared the odorant profiles of FKB to age-matched healthy pupae across six colonies located at three different apiaries. We found ten compounds that were consistently different between FKB and healthy pupae (Figure 3.3 A and B), although the identities of only four (isopropanol, 2-pentanone, β-ocimene and oleic acid) could be confirmed with commercial standards (Table 3.1). For the five unknowns (Compounds 1 to 5), either the retention times of the best matched (based on Mass Hunter scores) commercial standards did not match the peaks in the samples, making the identifications assigned by the spectral search algorithm unreliable, or the spectra could not be confidently matched to any in the comprehensive Wiley/NIST compound library. Of the ten compounds, nine were most abundant in the FKB headspace samples and only one was most abundant in live pupae. This peak had the highest volatility and a strong 44^+ base-peak ion,

which matches carbon dioxide and is consistent with active respiration of the live pupae. The carbon dioxide peak had above-background levels in the dead samples (although still significantly lower than in live samples), which is consistent with some decomposition beginning to occur at warm temperatures.

3.3.2 **Correlating odorant emission with colony hygienic score**

It has been established that hygienic adult workers have superior olfactory sensitivity compared to non-hygienic bees[91,92]; however, the brood itself could also play a role in the behaviour[202]. Since an odorant signal with a bigger difference between the healthy and diseased state should be easier for workers to discriminate and act upon, we hypothesized that brood from highly hygienic colonies may emit a stronger odour signal relative to healthy controls. In other words, there could be a brood effect contributing to overall hygiene. To test this, we correlated the dead:live ratio of each compound with hygienicity across eight different colonies. We found that the ratios of only one compound was significantly correlated with the behaviour: β-ocimene (Figure 3.4; Pearson coefficient = 0.84; $p = 0.0059$; $\alpha = 0.0063$; Bonferroni correction). β-ocimene is a familiar brood pheromone that is already known to increase worker visits to cells[206]; however, to our knowledge, it has not being previously associated with hygienic behaviour.

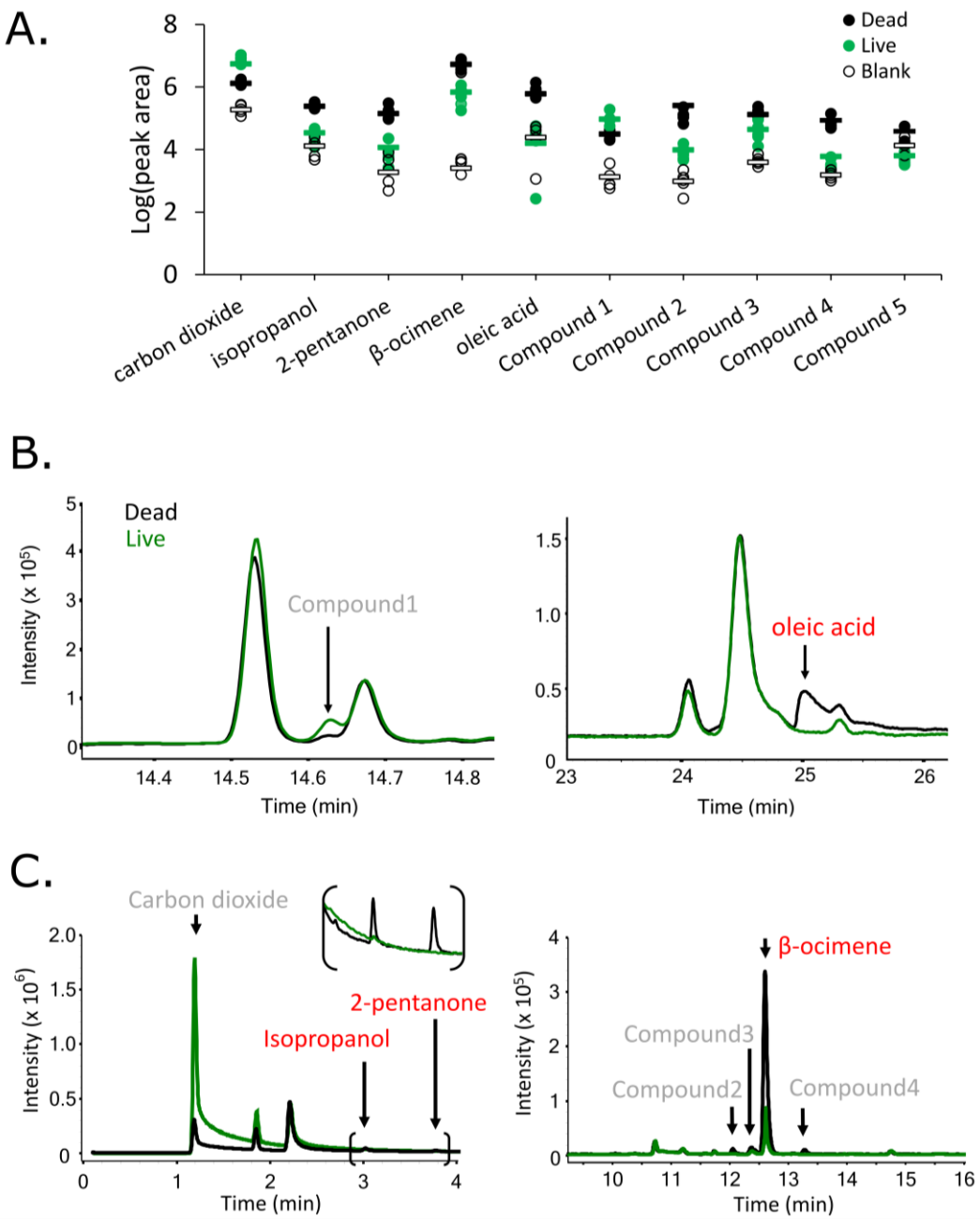


Figure 3.3. Cross-colony comparison of FKB and healthy brood odorant profiles.

A. Ten compounds were significantly differentially expressed across colonies (n = 6; two-factor ANOVA; Tukey HSD; see Table 3.1 for p values). Compounds 1 to 4 were identified as 2-methyl tetradecane, α-thujene, α-pinene and 2,3-butanediol, respectively, based on spectral matching to a database. Compound 5 (not displayed on chromatograms) could not be confidently matched to any spectra in the compound library. Bars represent averages. For cuticle compounds (the compounds in B), "Blank" represents clean solvent (hexane). For SPME compounds (the compounds in C) "Blank" represents an empty vial prepared alongside the samples. B. Cuticle hexane wash and C. solid phase micro-extraction (SPME) example chromatograms covering the differentially emitted compounds. Bracketed region is enlarged for clarity.

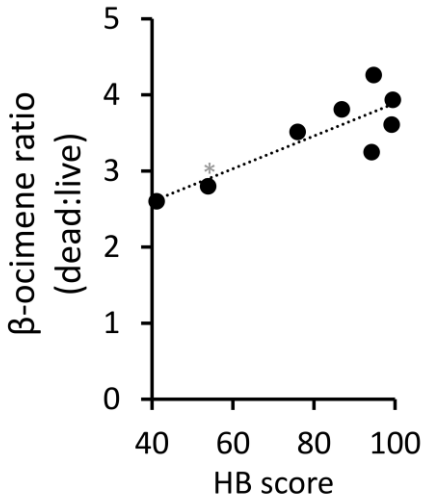


Figure 3.4. β-ocimene is a key compound in FKB, but not *Varroa*-infested brood.

Except where otherwise indicated, two rounds of hygienic testing (two patches in each round) were performed on eight different colonies. Out of all the significantly differentially emitted compounds, β-ocimene was the only one to significantly correlate with hygienic behaviour (Pearson correlation coefficient = 0.84, P = 0.0059, α = 0.0063; n = 3 pupae within each colony, averaged to produce the colony ratio; N = 8 colonies). *This colony was scored based on one round of hygienic testing.

3.4 Differential emission of odorants from *Varroa*-infested brood

To identify chemical cues associated with *Varroa* infestation, we compared odorant profiles between infested and non-infested brood. *Varroa* mites reproduce inside the developing pupa's comb cell, forming a whole family (including the foundress, eggs, protonymphs, deutonymphs and adult males) over time (Figure 3.1 C). We included four sequential developmental stages (white-eyed, pink-eyed, purple-eyed white body and purple-eyed tan body) and included the mite families with the pupae in the analysis. We did not find any significant effect of infestation on the headspace volatile profile (three-factor ANOVA; p = 0.46; F = 0.56); however, analyzing the cuticle profile showed that while infestation had no effect on its own (three factor ANOVA, p = 0.28, F = 1.15), it significantly interacts with developmental stage (Figure 3.5; p = 0.000022; F = 8.34). The overall trend was for infested brood to produce higher levels of cuticle compounds

relative to healthy brood in age-matched adjacent cells, but no individual compounds drove this effect.

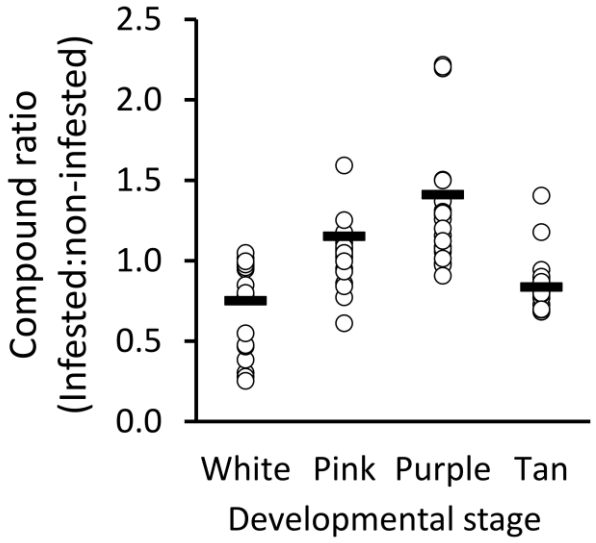


Figure 3.5. Effect of *Varroa* infestation on odorant emissions.

***Varroa*-infestation has a significant interacting effect (three-factor ANOVA; p = 0.000022; F = 8.34) on cuticle compound abundance, but specific compounds did not drive the effect (p = 0.99; F = 0.38).**

3.5 Electroantennographic recordings

We investigated the biological activity of isopropanol, 2-pentanone, β-ocimene and oleic acid using electroantennography (EAG) to quantify antennal nerve depolarizations of hygienic and non-hygienic bees in response to odorant stimuli (Figure 3.6). Of all the compounds, only 2-pentanone and β-ocimene showed dose-dependent responses (three-factor ANOVA; see Table 3.2 for summary statistics). For β-ocimene, we also found significant interactive effects between dose and hygienicity as well as hygienicity and antenna side. Notably, the left antenna of hygienic bees produced the strongest EAG signal overall – significantly higher than the right antennae – whereas non-hygienic bees did not display this effect. This is counterintuitive, since right antennae have a higher proportion of olfactory sensilla[207] and foragers are known to give

stronger EAG responses to (-)-linalool and isoamyl acetate (alarm pheromone) through their right antenna[208]. However, we confirmed that the same left-biased lateralization holds true for another known hygienic behaviour-inducing compound, phenethyl acetate[94] (isolated from chalkbrood; Figure 3.6 C; Table 3.2). Surprisingly, oleic acid appeared not to stimulate bee antennae at all, possibly because of its low volatility at room temperature. Isopropanol stimulations also produced no significant differences (Figure 3.7).

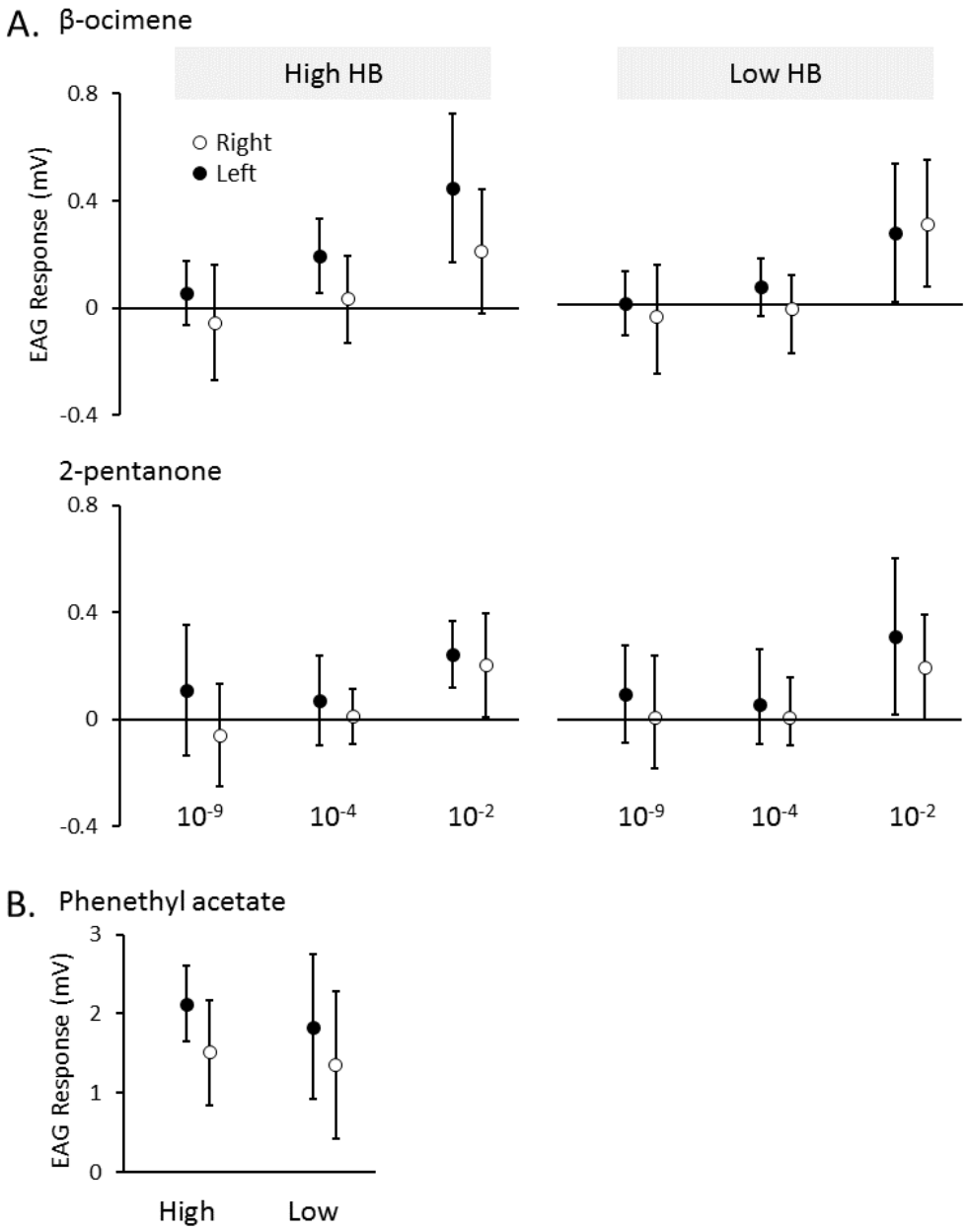


Figure 3.6. Antenna stimulations by candidate hygienic behaviour-inducing compounds.

Electroantennography (EAG) was used to quantify antenna responses to odorant stimuli. 2-pentanone and β-ocimene doses were applied at three dilutions (10^{-9}, 10^{-4} and 10^{-2} v/v). The response to 2-pentanone was dose-dependent and weakly lateralized but did not depend on hygienic behaviour (three-factor ANOVA). The response to β-ocimene was also dose-dependent and lateralized with a significant interactive effect between hygienic behaviour and dose as well as hygienic behaviour and side (three factor ANOVA). Phenethyl acetate, a known hygienic behaviour-inducing compound, was applied at one dose (10^{-9} v/v). A significant interactive effect was observed between hygienic behaviour and side (two-factor ANOVA). See Table 3.2 for statistical information and Table 3.3 for biological replicate numbers. Error bars represent standard deviation. HB = hygienic behaviour.

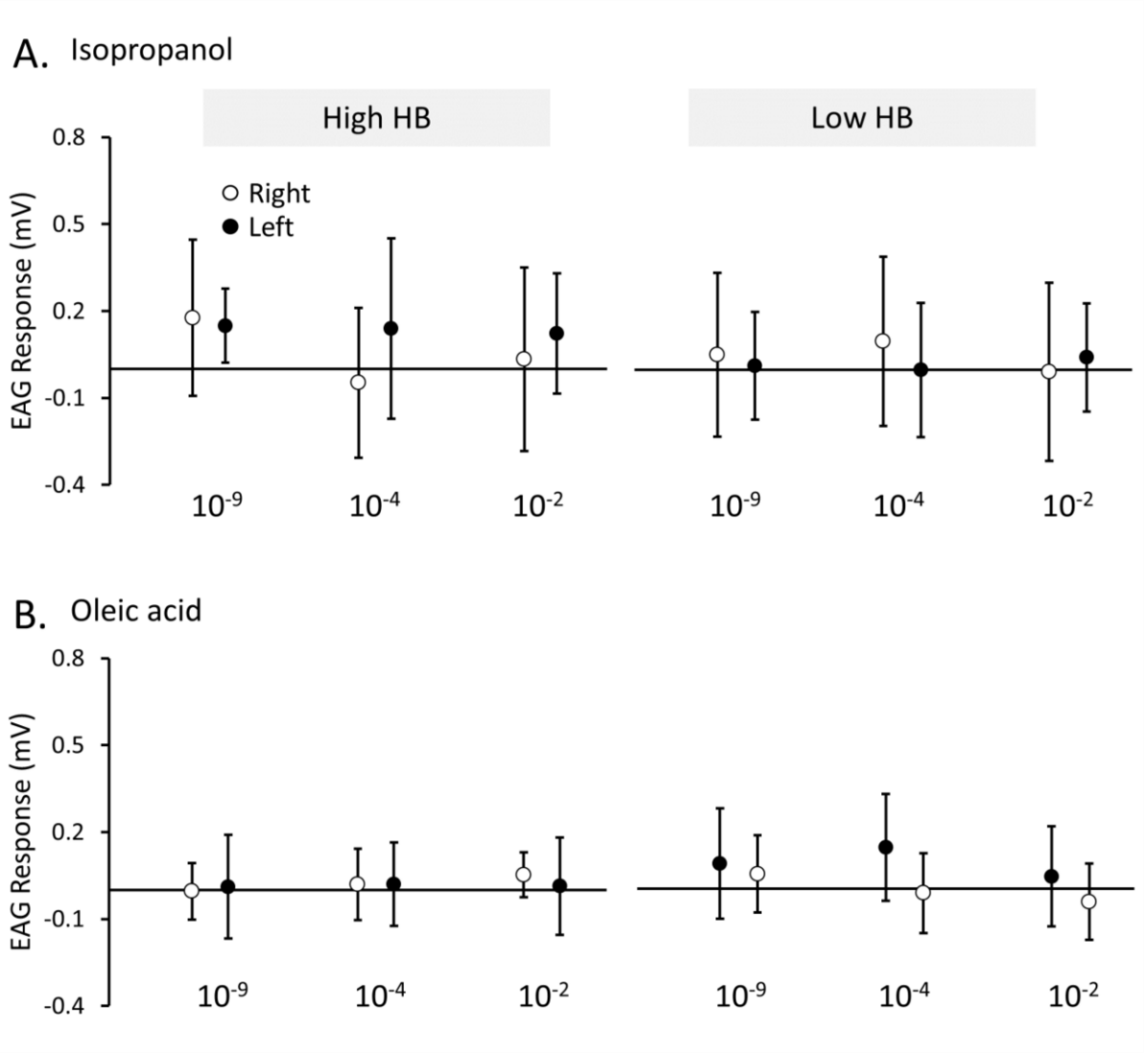


Figure 3.7. Electroantennographic data for oleic acid and isopropanol stimulations.

Doses were applied at three dilutions (10-9, 10-4 and 10-2 v/v in ethanol). No significant differences between dose, side or hygienic behaviour were found with these compounds (three-factor ANOVA).

Table 3.2. Summary statistics for EAG data

Compound	Groups	df	F	p-value	sig
Beta ocimene	Dose	2	33.4	1.20E-12	yes
	HB	1	1.9	1.68E-01	no
	Side	1	4.3	4.04E-02	yes
	Dose x HB	2	3.4	3.46E-02	yes
	Dose x Side	2	3.6	2.86E-02	yes
	HB x Side	1	4.8	3.08E-02	yes
	Dose x HB Behaviour x Side	2	0.6	5.42E-01	no
2-pentanone	Dose	2	17.3	1.73E-07	yes
	HB	1	0.3	5.74E-01	no
	Side	1	7.1	8.53E-03	yes
	Dose x HB	2	1.6	2.11E-01	no
	Dose x Side	2	1.3	2.70E-01	no
	HB x Side	1	1.5	2.24E-01	no
	Dose x HB Behaviour x Side	2	1.0	3.71E-01	no
Phenethyl acetate	Side	1	3.5	6.83E-02	no
	HB	1	1.5	2.28E-01	no
	Side x HB	1	36.2	2.73E-07	yes

Table 3.3. Replicates for EAG data

Dose	HB	Side	n
High	High	Left	10
High	High	Right	14
High	Low	Left	14
High	Low	Right	15
Low	High	Left	10
Low	High	Right	14
Low	Low	Left	14
Low	Low	Right	15
Med	High	Left	10
Med	High	Right	14
Med	Low	Left	14
Med	Low	Right	15

A well-known phenomenon in olfactory perception is the synergistic effect of odorant mixtures[209]. That is, mixtures can sometimes be perceived not as the sum of their parts, but as if they are entirely new odours; however, this is rarely observed in honey bees[210-212]. To test if the four odours could lead to stronger EAG signals by stimulating antennae synergistically, we produced equivolume mixtures (1% total in ethanol) of all possible combinations of isopropanol, 2-pentanone, β-ocimene and oleic acid and used these to perform EAG on left antennae of hygienic bees. None of the odorant combinations induced greater antenna stimulations than β-ocimene alone (the strongest stimulator; Figure 3.8).

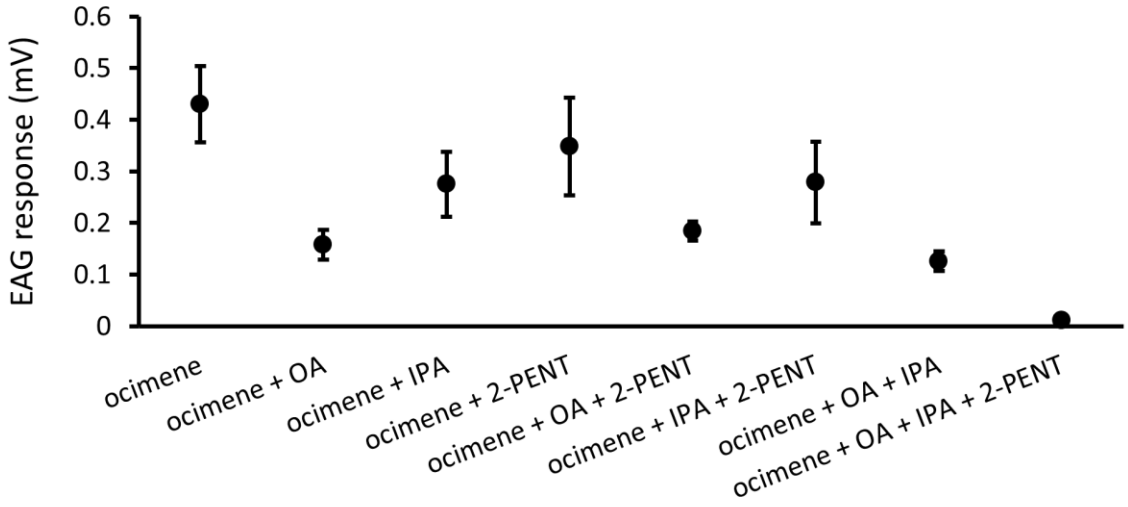


Figure 3.8. Combinatorial analysis of disease odors.

All combinations of oleic acid (OA), isopropanol (IPA), 2-pentanone (2-PENT) and β-ocimene were mixed to 1% solutions (v/v) and used to stimulate honey bee worker antennae. No stimulations produced stronger EAG responses than β-ocimene (mixtures without β-ocimene are not shown). Error bars represent standard deviation.

3.5.1 **Doubts raised over lateralization**

Previously, we used electroantennography to compare the responses of left and right antennae from hygienic and non-hygienic bees upon stimulation with β-ocimene. In that test, we found that left antennae responded more strongly to the odorant than right antennae, but we questioned

this result because it is not consistent with most patterns in the literature (typically, the right antenna is more sensitive than the left). Here, we repeated the comparison of the left and right antenna response to β-ocimene using a larger sample (N = 22) of bees from a single highly hygienic colony. We observed the opposite lateralization pattern to what we measured before (Figure 3.9 A). We still observed a significant dose-response (two-way ANOVA; $F = 27.5$, $p = 1.5e-10$), but this time, the right antennae responded significantly more strongly than the left ($F = 6.3$, $p = 0.01$).

In all our tests, β-ocimene is diluted in ethanol, then during data analysis the response to ethanol alone is subtracted to yield the EAG response to β-ocimene alone. In order to determine if our results could be explained by differences in the response to ethanol alone, we examined the ethanol responses in our previous experiments as well as the present one (Figure 3.9 B). We found that in our previous experiment, for hygienic bees, the right antenna tended to produce a higher EAG response to ethanol compared to the left (although this was not significant; three-factor ANOVA; Tukey HSD post hoc test; $p = 0.20$), whereas in the present study, the left and right are more similar ($p = 0.87$). For a given β-ocimene (diluted in ethanol) stimulation, then, subtracting a background stimulation that has a larger magnitude will yield a lower EAG value for β-ocimene alone. Furthermore, we also found that the right antenna of hygienic bees responds significantly stronger to ethanol than the right antenna of non-hygienic bees, which could be the real basis to the apparent hygienic-associated lateralization. Together, these patterns are likely what caused the previously observed left-biased β-ocimene EAG responses for hygienic bees, which is absent in the present study. It is still unclear why the response to ethanol could be so

variable, and this has prompted us to interpret any apparent lateralization in EAG signal with great caution in the future.

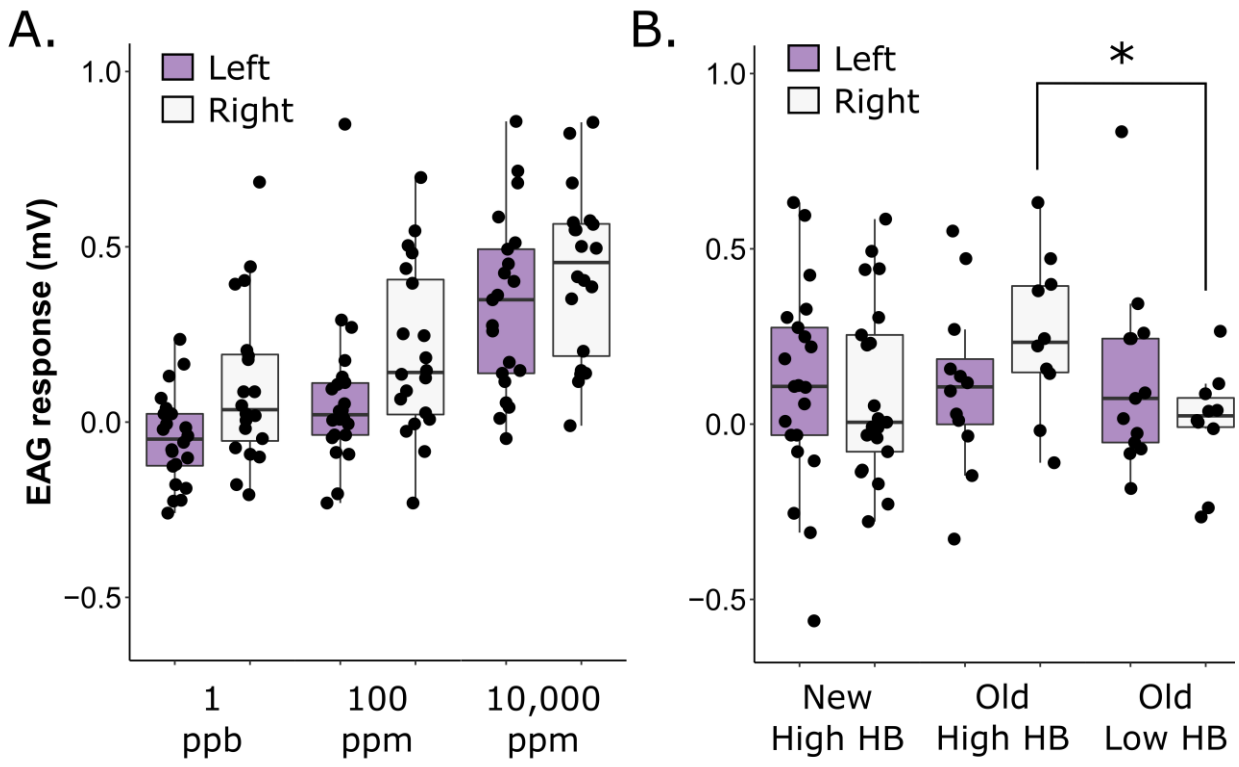


Figure 3.9. Lateralization follow-up EAG recordings.

A. Bees' antennae from a single hygienic colony (hygienic score: 95%) were excised and stimulated with three increasing concentrations of β-ocimene (N = 22). The plotted EAG response is the background (ethanol)-subtracted EAG amplitude. Data was analyzed with a two-factor ANOVA (levels: side and dose). There was a significant effect of dose (F = 27.5, p = 1.5e-10) and side (F = 6.3, p = 0.01). Boxes depict the interquartile range (IQR), whiskers span 1.5*IQR, and bold bars represent the median. B. EAG responses to ethanol alone. Plotted values are blank-subtracted ethanol stimulations (i.e. the ethanol EAG response minus the response to air alone). "New" refers to the present study, whereas "Old" refers to the previously published study [3]. There was a significant interactive effect between hygienicity and side (three factor ANOVA; F = 5.02; p = 0.035). A Tukey HSD post hoc test revealed this was driven by the contrast indicated by an asterisk (p = 0.0314).

3.6 Proteomic investigation of antennal lateralization

Whether antenna lateralization is left-biased or right-biased, we were interested in determining a potential mechanism for lateralization at the gene expression level because numerous other publications report lateralization having a role in other contexts for honey bees and other bee species[207,208,213-219]. To investigate this, we performed label-free quantitative proteomics on left and right antennae of nurse bees from five hygienic colonies. Despite identifying 1,845 proteins (13,128 peptides) at 1% FDR, none of them were differentially expressed. Interestingly, 230 of the identified proteins are ones that were discarded from the first Official Gene Set (OGSv1.0), apparently in error, which we had added back to our protein database after our protegenomics analysis Chapter 2, Section 2.4.2). We described this phenomenon previously[151] and this finding offers secondary confirmation. A further 15 proteins are new, previously unannotated candidate protein sequences which we identified in the same previous proteogenomic effort.

We reasoned that one explanation for failing to find differentially expressed proteins, despite the left and right antennae being known to have different distributions of sensilla and lateralized function, could be because the depth of coverage we achieved was insufficient to observe differences that may exist in low-abundance proteins. To improve proteome coverage in the current study, we fractionated the peptides from left and right antennae from 4 highly hygienic colonies and repeated the comparison (Figure 3.10). We increased proteome coverage from 1,845 to 3,114 unique proteins, which is among the highest proteome coverage achieved in honey bees to date[151]. However, we still did not identify significant differences. Hierarchical clustering shows that left and right antennae of bees from the same colony cluster more closely than antenna sides across colonies.

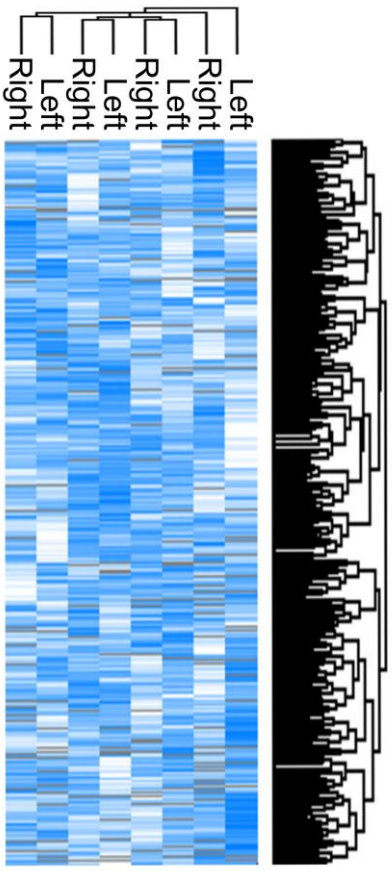


Figure 3.10. Comparing the proteomes of left and right antennae from hygienic honey bees.

Only the data from the deep proteomics analysis is shown. Digested peptides were analyzed on a Bruker Impact II Q-TOF mass spectrometer and label-free quantitation (LFQ) was used to compare protein expression between samples. 3,114 proteins were identified, but no proteins were significantly different at 10% FDR (Benjamini-Hochberg correction). Z-score scale: white = -2.5, blue = +2.5, and grey = not quantified. Hierarchical clustering was performed in Perseus using average Euclidian distance (300 clusters, maximum 10 iterations).

3.7 Conclusion and future directions

Overall, our experimental findings point to emerging mechanistic patterns of hygienic behaviour. We found that a well-known brood pheromone, β-ocimene, was strongly emitted from FKB and this pattern positively correlates with hygienic behaviour score. We also identified one compound, oleic acid, which was consistently released in higher amounts in FKB, not only

across colonies but also across developmental stages. Finally, we functionally validated these compounds using electroantennography and while we found initially that lateralization of antennal response is strongly associated with hygienic behaviour, upon closer inspection we found that this was instead most likely a result of aberrant lateralized response to the solvent, rather than the odorant of interest. In addition, we could not identify proteomic changes associated with lateralization, despite producing one of the deepest honey bee proteomics datasets to date. Unlike the FKB, we found that *Varroa*-infestation causes subtle but significant changes to the overall cuticle compound profile, although no individual compounds emerged as drivers. This may in part explain why trait selection for *Varroa*-sensitive hygiene, a specialized form of hygienic behaviour, requires more rigorous selection techniques.

3.7.1 **Odorants emitted from freeze-killed brood**

The clear majority of differentially emitted compounds were more abundant in dead pupae than in live ones. This is intuitive, since hygienic behaviour-triggering compounds should give a more reliable and specific signal to the bees if dead:live discrimination is based on their presence, rather than absence. Furthermore, adding extracts from dead or diseased brood to live brood is known to induce hygienic behaviour, which would not occur if the signal was based on a compound's absence[47,91,92]. Interestingly, three different putative terpenes were identified (based on a fragment base peak of 93.0 m/z and spectral matches to other terpenes; Table 3.1) but of these, only β-ocimene could be confidently confirmed. These unidentified compounds could still certainly be biologically relevant to hygienic behaviour, but further work is required to identify them.

Oleic acid – an omega-9 monounsaturated fatty acid – was emitted more strongly in FKB compared to live brood of all ages tested in this study. Intriguingly, oleic acid has been implicated as a mechanistic agent for hygienic behaviour in other social insects[64,66]. For example, Wilson *et al.*[64] found that applying oleic acid to otherwise mobile and healthy ants induced other ants to transport them to their refuse area. Oleic acid also binds strongly to odorant binding protein 18, which significantly correlates with hygienic behaviour in honey bees[199] and is currently being used for marker-assisted selective hygienic behaviour breeding[37]. Finally, it has been shown previously that *Varroa*-parasitized brood (which can also trigger hygienic behaviour) may emit more oleic acid compared to healthy brood; however, at the time of that study it was not identified as a discriminating compound[203], and we could not replicate these results in our analyses of *Varroa*-infested pupae. Oleic acid is also a component of the *Varroa* sex pheromone[173]. Despite these lines of evidence, oleic acid did not elicit a strong EAG signal, possibly because of its low volatility. Oleic acid's boiling point is 360°C, whereas 2-pentanone, the next highest, boils at 101ºC.

β-ocimene is a well-known brood pheromone that plays multiple roles in regulating worker behaviour and anatomy[206,220-224]. Young larvae normally release β-ocimene to stimulate workers to feed them[206], with levels tapering off with age. β-ocimene also plays a role in regulating forager activity[224] and inhibits worker ovary development[222], but, to our knowledge, its release has not been previously associated with the FKB assay or hygienic behaviour. It seems unusual that dead bees would emit more of a brood pheromone than live bees, but it could be that a normally tightly controlled pheromone release mechanism breaks down as membranes become more permeable after freezing. An independent study found that a different brood pheromone

(brood ester pheromone; BEP) was also significantly more abundant in *Varroa*-parasitized brood[96]. By increasing worker visits to brood cells that should otherwise not require attendance, β-ocimene may attract the attention of bees that can perform hygienic behaviour. As the second brood pheromone implicated in hygienic behaviour, these results may indicate a broader pattern of hygienic behaviour dependence on brood pheromones. Furthermore, Mondet *et al.*[96] suggest that BEP contributes to detection of *Varroa*-infested pupae by signaling developmental delay – this mechanism is consistent with our own observations, since more β-ocimene is emitted from young larvae compared to pupae[221].

Further stimulating our interest in this compound, we also found that the ratio of β-ocimene emitted from FKB to live pupae significantly positively correlates with hygienic behaviour itself. This suggests that there may be a brood effect contributing to hygienic behaviour scores, in addition to olfactory sensitivity of adult workers, even though previously this was not thought to be the case. In an early foundational paper, Spivak and Downey[47] found no consistent brood effect when they performed hygienic tests using reciprocally donated brood; however, brood age was not controlled during these tests, and introducing hygienic brood to non-hygienic colonies was not investigated. In other words, the hygienic score of the donor was not a factor in the analysis, as far as we can tell. In the same study, they established that from year to year, brood age had a significant effect on non-hygienic colonies but not on hygienic colonies, with non-hygienic colonies performing significantly better on the FKB test when young brood (capped larvae and prepupae) was used compared to older pupae. However, this effect was not observed in tests performed the following year. The confounding factors may have simply erased potential

brood effect patterns. Since our analysis only evaluated eight colonies, and only during one season, this potential brood effect warrants further investigation across more colonies.

This is not the first time that a brood effect has been suggested as a component of social immunity. In VSH, it is a well-established observation[38,85,89]. For example, Parker *et al.* (2012)[38] found significant differences in the larval cuticle proteome between high and low VSH bees and suggested that this may lead the brood to emit different chemical cues. This data, together with our own, suggests that hygienic behaviour could also be dependent on two interacting factors – the strength of brood odour and the workers' limit of odour detection – rather than the adult workers' olfactory sensitivity alone. It would be worthwhile to examine this effect across more diverse sources of high and low hygienic behaviour colonies to determine if it is a ubiquitous theme, or if there is a distribution of colonies that achieve hygiene primarily through a brood effect, worker olfactory sensitivity, or a mixture of the two.

3.7.2 **Odorants emitted from *Varroa*-infested brood**

When we compared odour profiles of *Varroa*-parasitized pupae to healthy pupae across four developmental stages, we found a significant interaction between parasitization and developmental stage but no individual compounds drove this effect (Figure 3.5). This could be because VSH is a specialized form of hygienic behaviour[84] and this specialization may be in part required because the differences between infested and non-infested brood are subtler than for dead and live brood and require either ultra-sensitive olfaction or a contributing brood effect to amplify the odorant signal. Indeed, one proposed strategy for mites to evade detection in the colony is to adapt their own cuticle hydrocarbon profile to mimic its host[225]. Another explanation

is that since the healthy control brood was pulled from cells immediately adjacent to the infested pupa, it could be that *Varroa*-associated compounds transferred through the thin wax wall to the healthy pupae, diminishing the observable differences. However, we think contamination of this magnitude is unlikely and we still believe that this was the appropriate comparison, since hygienic bees must be able to discriminate between neighbouring healthy and diseased states. Finally, it could also be that key differentially emitted compounds do exist, but we were unable to detect them with our extraction method or our sample size. On the other hand, it could be that *Varroa* infestation really does cause little, if any, differential compound emission due to selective pressure to evade detection by the host.

Mondet *et al.*[96] were recently able to find *Varroa*-specific compounds by analyzing solvent extracts of crushed infested pupae. However, it is not clear that compounds measured in this way would be detectable by bees performing hygienic behaviour, and the odorant cues were not confirmed behaviourally. The hexane extraction and SPME used here are suitable for capturing non-polar compounds with relatively high volatility but it could be that the superior olfactory sensitivity of hygienic bees allows them to detect some polar, non-volatile compounds. Indeed, oleic acid (a carboxylic acid) is one of our most confident hygienic behaviour-inducing candidates but it was among the last to elute in our GC-MS analysis of hexane extracts; more polar compounds would likely become trapped in the GC-MS inlet or not be miscible in hexane at all. Notably, Mondet *et al.* also observed that P5 pupae (roughly equivalent to our purple-eyed white-body stage) are targeted most frequently for VSH, following the same trend as overall compound abundance in infested pupae as displayed in Figure 3.5). This points to the possibility that VSH bees are either a) more sensitive to a specific compound found in the milieu of more

abundant compounds emitted from *Varroa*-infested brood or b) VSH bees are broadly more sensitive to a suite of compounds associated with infestation.

3.7.3 **Lateralization of olfactory sensitivity**

The left-biased EAG response lateralization was intriguing, but based on our follow-up experiments, we interpret this result extremely cautiously. Lateralization in bees is not new – for example, Rogers *et al.*[218] have shown that bees are more likely to interact aggressively when communicating *via* their left antenna, whereas they have preferentially positive encounters when interacting *via* their right antenna. Furthermore, Rogers and Vallortigara[217] found that bees performed better at long term memory recall tasks when stimulated *via* their left antennae, but not their right. We did not acquire data on the higher order processing of odours, but these studies create a precedent for antenna lateralization as it relates to behaviour. However, we thought it was unusual that despite having more olfactory sensilla on the right antenna[207], the left elicited a stronger EAG signal for FKB and chalkbrood compounds. One possible explanation is that the olfactory sensilla that do exist on the left antenna house olfactory receptor neurons that are specifically tuned to particular odours, but based on our subsequent experiments, we think it was more likely to be caused by a fluke of right-biased responses to the background solvent (ethanol). These experiments highlight the importance of replication not only within a study, but between studies.

3.7.4 **Conclusion**

The work presented here furthers our understanding of hygienic behaviour and the underlying mechanism. Interestingly, this is now the second study to implicate a previously known brood

pheromone (β-ocimene, in this case) as a potential mechanistic agent for hygienic behaviour. We found that bees' antennae responded in a dose-dependent fashion to β-ocimene and 2-pentanone, but not oleic acid or isopropanol. However, since it is already known that odorant binding protein 18, which is thought to aid in odour detection, positively correlates with hygienic behaviour and that oleic acid is one of its strongest known ligands, we suspect that the failure of antennae to respond to this compound is due to its low volatility. Oleic acid is also known to induce hygienic behaviour in other social insects, suggesting that the mechanism for hygienic behaviour could be evolutionarily conserved. The odour profiles of *Varroa*-infested brood showed a significant interaction between infestation and developmental stage, but no individual compounds were significantly different. This subtlety is consistent with VSH being a specialized form of hygienic behaviour, which may be tuned to specific odorants below our limit of detection or more general signals like developmental delay, as suggested by Mondet *et al*[96]. Since we did not use internal standards in the GC-MS analyses, these experiments offer only a preliminary screen for differentially emitted compounds based on rough relative quantitation. Further experiments are needed to confirm the identities of the five unknown significant compounds, since they may still be biologically relevant to hygienic behaviour, and to confirm that the candidate compounds really induce hygienic behaviour in a realistic setting.

3.8 Methods

3.8.1 Data access

Mass spectrometry raw data for the initial (low-depth) lateralization comparison is available at Pride ProteomeXchange (www.proteomexchange.org; PXD005242). The mass spectrometry data

for the high-depth lateralization comparison are available at ftp://massive.ucsd.edu/MSV000081790.

3.8.2 Sample collection

3.8.2.1 Freeze-killed brood

Honey bee pupae with no visible signs of disease were collected from colonies by carefully uncapping cells and removing pupae with clean stainless-steel forceps. Age was determined based on eye and cuticle pigment using the following relationships: white-eyed = 12-13 d, pink-eyed = 14-15 d, purple-eyed white body = 16 d and purple-eyed tan body = 17-18 d. From bee to bee, eye and cuticle pigment was matched exactly so that each bee in each age group was at the same developmental stage. Pupae were placed in clean glass vials, removing any wax debris and avoiding abrasions or cuticle indentations. Freeze-killed samples were placed at -80°C (15 min) then placed in a humid 33°C incubator (24 h), while live samples were placed directly into the same incubator. After the 15 min freeze, all pupae were completely solid and brittle so there is no doubt that they were mortally frozen.

Compounds were extracted for low resolution GC-MS analysis by two different methods: solvent extraction and solid phase micro-extraction (SPME). For analyzing cuticular compounds across developmental stages (white-eyed, pink-eyed, purple-eyed white body and purple-eyed tan body; n = 3), extracts were prepared by washing whole pupae with 300 µl HPLC-grade hexane for 5 min with gentle agitation. Hexane extracts were transferred to a clean vial and immediately stored at -80°C until GC-MS analysis. For the cross-colony analysis (N = 3 per colony, n = 6 colonies), compounds were extracted only from purple-eyed white body pupae using the method

above as well as by sealing individual freeze-killed and live pupae in 10 mL glass vials (Supelco) and incubating at 33°C (24 h) prior to SPME analysis. We confirmed that 10 mL of air is enough for one bee to survive for this time by performing the same procedure for late-stage pupae, which were still actively moving after being sealed for 24 h.

One µL of each hexane extract was analyzed by GC-MS (Agilent 6890N/5975C Inert XL MSD) using a DB-wax column (J&W 122-7032) and a 30 min gradient from 50°C to 230°C. The back inlet (pulsed splitless) was at 250°C and 6.24 psi with a 53.5 mL/min flow rate (He gas) connected to the analytical column (30 m, 250 µm ID). The instrument was set to scan from 40 to 300 m/z. The MS source and quadrupole were maintained at 230°C and 150°C, respectively. Headspace volatiles were sampled using solid phase micro-extraction (SPME) and analyzed by GC-MS (Agilent 7890A/5975C Inert XL MSD) using a 45 min gradient and the same column model as above. We used a 50/30 µm DVB/CAR/PDMS stableflex SPME fiber (Sigma) and sampling details were: 40°C incubation, 3 s agitation at 500 rpm, 600 s extraction time and 300 s desorption time. The oven settings were: 35°C (stable; 4 min), then 25°C/min (5 min) and a 2:1 split ratio. The inlet temperature was 250°C and MS acquisition parameters were the same as above except that the lower mass limit was 33 m/z.

3.8.2.2 *Varroa*-infested brood

For ease of sampling, mite-infested brood were concentrated on a single frame by caging the queen in a single-frame excluder and transplanting all other open brood into a temporary 'incubator' colony. This left only the single frame of brood suitable for mite infestation, effectively concentrating the phoretic mites looking for brood cells in that colony to one location.

After 10 d, the brood was returned from the incubator colony and the queen was released. Following this, pupae were sampled by the same methods as above and only pupae with a single foundress mite were chosen. The accompanying mite family (including foundresses, deutonymphs, protonymphs and eggs) was transferred to the same glass vial as the pupa using a soft paintbrush. Adjacent, age-matched non-infested sister pupae with no visible signs of disease were collected as healthy controls.

3.8.3 **Hygienic testing**

Honey bee colonies were kept at three separate locations in Greater Vancouver, BC, Canada. Colonies were scored for hygienic behaviour using the FKB assay as previously described[47]. Briefly, for each test, polyvinyl chloride pipes (5 cm inner diameter, ~25 cm length) were pressed into capped brood comb in two areas containing white-eyed to red-eyed pupae, then filled with approximately 250 ml of liquid nitrogen to freeze. Frames were returned to the colony and assessed 24 h later for percent removal of the frozen brood cells. One week later, the test was repeated, and the average of the two tests (four 5 cm brood patches in total) yielded the FKB score. All testing and sampling was conducted during the summer of 2016.

3.8.4 **Gas chromatography mass spectrometry data analysis**

GC-MS data was analyzed using Mass Hunter Qualitative Analysis software (vB.06.00). Chromatogram peaks were first smoothed using the default algorithm and then manually integrated to ensure consistent baselines between replicates. Internal standards and calibration curves were not utilized in this analysis; rather, peak areas were compared directly. This method of relative quantification is not as accurate, but it offers a quick screen for candidate

differentially emitted compounds as long as the column does not reach saturation. To perform relative quantification of odorant profiles of FKB to healthy pupae across developmental stages, peak areas were exported to Excel (2013) where they were $\log_{10}$ transformed and groups (developmental stage, freezing, compound type) were compared using three-factor ANOVA (Excel), followed by a Tukey HSD *post-hoc* test to identify the specific differentially emitted compounds. We did not test the data for normality, but the ANOVA is generally tolerant to non-normal data and/or low replication. The same process was used to analyze FKB changes across colonies except that a two-factor ANOVA was employed (since this only involved a single developmental stage: purple-eyed white body pupae). The effect of *Varroa*-infestation was also examined using a three-factor ANOVA (developmental stage, infestation, compound). In all cases, compound identities were determined by searching spectra against the Wiley Chemical Compound Library (W9N08.L) in Mass Hunter, but only some compounds of interest (where indicated) were confirmed against a commercial standard.

To determine if any of the significantly differentially emitted compounds – including those with unassigned identities – correlated with colony hygienic score, we calculated the dead:live ratio (not log transformed), then the Pearson correlation for each compound. We did not attempt to correlate the carbon dioxide abundance, since this is simply a result of respiration, nor did we include compound 5, which was significantly higher in the FKB compared to live, but not to the empty background. In total, 8 compounds were correlated. To account for multiple hypothesis testing, significance was determined by comparing p-values against the Bonferonni-corrected α ($0.05/8 = 0.0063$).

3.8.5 **Antenna preparation for electroantennography**

Bees for electroantennography (EAG) were collected across three colonies with high hygienic behaviour scores and three with low hygienic behaviour scores. Since bees perform hygienic behaviour best when they are two to three weeks old[226], we marked emerging bees with a paint pen and returned them to the hive for 14 d, then EAG data was acquired for up to one week. Antennae were excised and both ends were trimmed with a scalpel, randomizing whether right or left antennae were excised first. Trimmed antennae were then attached to glass capillary reference and recording electrodes filled with insect saline solution (210mM NaCl, 3.1mM KCl, 10mM $CaCl_2$, 2.1mM Na_2CO_3, 0.1 NaH_2PO_4) as previously described[227]. EAG responses were recorded on the EAD program of a Syntech™ IDAC-4 signal acquisition unit. The low cutoff was set at 0.1 Hz, high cutoff at 10 Hz, external amplifier set to 1. Humidified, charcoal filtered air was passed continuously over the antenna via a Syntech CS-55 stimulus controller, also serving as a carrier for odour-filled pulses. Odorants were dispensed onto 1 cm^2 No. 1 Whatman filter paper, allowing the solvent to evaporate for 30 s before being inserted into a glass Pasteur pipette. Odorant pulses were passed through the Pasteur pipette to the antenna for 1 s, and between 0.5 and 1 min was allowed between each presentation of an odour for the antenna to return to baseline activity. Each antenna was stimulated with a series of three dilutions (10^{-9}, 10^{-4} and 10^{-2} v/v in ethanol) each of isopropanol, 2-pentanone, β-ocimene and oleic acid (all from Sigma or Fisher; >90% purity). Phenethyl acetate, a known hygienic behaviour-inducing compound isolated from chalkbrood[94], was used as a positive control. All possible equivolume combinations of the four candidate compounds were also tested at a 10^{-2} (1%) dilution to test for synergistic effects of mixtures.

Even though antennae were conditioned with humidified air throughout the recordings, EAG signal decay was still evident even for stimuli of solvent alone over time. Therefore, each antenna was subject to intermittent solvent stimulations throughout the recordings to mathematically interpolate the background solvent stimulus. The quality cut-off for the solvent curve fit was $R^2 > 0.8$: traces which did not meet this criterion were discarded. The final replicate numbers included in subsequent analyses are show in Table 3. Since the number of surviving traces varied, in total we acquired between 10 and 15 biological replicates in each experimental group (left vs. right; high hygienic behaviour vs. low hygienic behaviour). Finally, the interpolated solvent amplitude was then subtracted from the solvent + odour stimulations, resulting in the mV value that can be attributed to the odorant alone. The amplitudes of our recordings are consistent with other similar studies in bees[91,93,217]. Statistical analyses were conducted using a three-factor ANOVA (dose, hygienic behaviour and antenna side) and a Tukey HSD post-hoc test.

3.8.6 **Protein extraction and processing**

Thirty to forty bees on open brood frames were collected from four highly hygienic colonies (n = 5; all with FKB scores > 94%). Bees were anesthetized with carbon dioxide and their antennae dissected on ice followed by homogenization (Precellys 24; Bertin instruments) with ceramic beads (lysis buffer: 6M guanidinium chloride with 10 mM TCEP, 100 mM Tris (pH 8.5), 40 mM chloroacetamide). The homogenizer was set to 6,400 rpm for 30 s x 3 (1 min on ice in between). Lysate was transferred to a new tube and debris was pelleted (14,100 rcf, 15 min, 4°C), followed by acetone precipitation as previously described[228]. Dried protein pellets were resuspended in 50 mM ammonium bicarbonate buffer (1% sodium deoxycholate) and protein concentration was

determined using a bicinchoninic assay (Pierce). Protein was reduced, alkylated, digested and analyzed on an LC-ESI-MSMS system (Easy nLC-1000 coupled to a Bruker Impact II mass spectrometer) as described in our previous publication[151], except we loaded 2.5 µg (based on protein quantitation), the LC gradient was 165 min and MSMS frequency was set to 18 Hz (see embedded microTOFQImpactAcquisition.method files within PXD005242 for further details).

For the deep antennal proteomics experiment, we dissected approximately 30 pairs of worker bee antennae from each colony (N = 4 hygienic colonies). Proteins were extracted and processed for mass spectrometry exactly as described above, except after digesting 30 µg of protein, the peptides were fractionated using basic reverse phase chromatography[229]. We pooled every 6th fraction, dried them down (Eppendorf Speed Vac), and acidified them in 0.5% formic acid prior to loading 20% of the sample (approximately 1 µg) on a Bruker Impact II Q-TOF mass spectrometer (coupled to a Thermo EASY-nLC 1000 chromatography system) for shotgun proteomics analysis.

3.8.7 **Proteomics data analysis**

Proteomics data was searched using MaxQuant (v1.5.5.30) and processed using Perseus (v1.5.5.3). All MaxQuant search parameters were left as default except: deamidation (NQ) was added as a variable modification, “match between runs,” “label-free quantification” and “re-quantification” options were enabled and “min ratio count” was set to 1. Briefly, reverse hits, proteins “only identified by site” and contaminants were removed followed by filtering for proteins identified in four or more colonies. Data was then Log2 transformed and missing values

were imputed before comparing left and right antennae using a t-test (Benjamini-Hochberg corrected 10% FDR).

Chapter 4: Behavioural validation of hygienic behaviour-inducing odorants[4]

4.1 Introduction and rationale

Previously, we compared odorant profiles of freeze-killed pupae and healthy pupae to find candidate hygienic behaviour-inducing compounds[230]. Although freeze-killing is not a natural cause of death, it is a relevant system because the freeze-killed brood assay[47] is the main method for determining colonies' level of hygiene. We identified several candidate hygienic behaviour-inducing compounds, two of which we will focus on here: oleic acid and β-ocimene. Oleic acid is a non-volatile, oily substance which acts as a death cue in eusocial and non-eusocial insects[66,69,70,78,97-102]. β-ocimene, on the other hand, is a volatile honey bee brood pheromone that, among other functions, is normally a larval food-begging signal[206]. β-ocimene and oleic acid have not been previously linked to hygienic behaviour in honey bees, and the ability of these compounds to induce hygienic behaviour in realistic bioassays has not been investigated.

4.2 Research goals

In the present work, our goal is to investigate oleic acid and β-ocimene's potential roles in hygienic behaviour. Previously, we found that oleic acid did not stimulate antennae during EAG recordings, despite it being a conserved necrophoretic compound. However, these EAG tests were not performed at hive-realistic temperatures; therefore, we repeated these measurements at ~35ºC to better understand how such a viscous compound could be detected in a sealed brood cell. Next, we use a series of behavioural assays to test the odorants' abilities to induce hygienic

[4] Content in this chapter has been published: McAfee, A. *et al.* A death pheromone, oleic acid, triggers hygienic behaviour in honey bees (*Apis mellifera* L.). *Sci. Rep.* 8:5719. Copyright (2018) Authors. Permission not required for reprinting. This work is licensed under a Creative Commons Attribution 4.0 International License.

behaviour in a realistic, in-hive setting: 1) the front-way odorant assay, where pupal cells are uncapped and odorants are dispensed on their heads, 2) the back-way odorant assay, where Jenter cages are used to introduce odorants without disrupting the cap, and 3) the broodless odorant assay, which is the same as in 2), but the brood has been extracted and replaced with an odorant-filled piece of filter paper. The Jenter approach[47,231] overcomes a major hurdle in testing the hygienic behaviour-inducing capacity of odorants: by adding odorants through the back of specially engineered, resealable brood comb, individual odorants are added to brood cells while maintaining perfect integrity of the wax cell walls and cap.

4.3 Front-way odorant assay

To test if β-ocimene and oleic acid are sufficient to induce brood removal, we used a front-way odorant assay (Figure 4.1 A), which involves uncapping patches of brood (30 cells each, in two technical replicates per colony) and dispensing 1 µl of either neat (100%) or diluted (1%) odorant standards on the brood.

4.3.1 Preliminary tests for developing the front-way odorant assay

First, we confirmed that hexane was an appropriate negative control by recording the recapping frequencies following the treatments (N = 9 colonies). We found that after just three hours, an average of 44% of the hexane-treated cells were recapped, which was significantly higher than for all other odorants (Figure 4.1 B**;** one-way ANOVA followed by Tukey HSD; hexane compared to β-ocimene: p = 2e-7; oleic acid: p = 1e-8; mix: p = 1e-8; phenethyl acetate: p = 1e-8). The next highest was β-ocimene, with 5.4% recapped. The others all had recapping frequencies of 1% or less, indicating that the brood were no longer accepted by the workers.

We also investigated the contact toxicity of the odorants themselves in an acute toxicity assay (Table 4.1). We found that 1 µl of 1% odorants induced almost no visible toxicity after 1 d (scored by the presence or absence of a prophenoloxidase response, or melanization); however, neat odorants did cause considerable toxicity, with β-ocimene inducing the most (40%, or 6 out of 15 pupae). We performed another toxicity assay, this time incubating older (purple-eyed, white body) pupae for 2.5 d. In this assay, oleic acid induced the most toxicity (also 40%). Overall, the toxicity patterns are not consistent and are difficult to interpret. A more reliable and robust toxicity assay should be developed in the future.

Table 4.1. Odorant toxicity assays

	1 d, white-eyed pupae		**2.5 d, purple-eyed pupae**
Odorant	**1% odorants**	**100% odorants**	**100% odorants**
Hexane (all 100%)	0[1]		0
β-ocimene	0	40	0
Oleic acid	0	6.7	40
Mix	6.7	6.7	27

[1]15 pupae were treated in each group

Next, we sought to confirm that there was no effect of patch proximity on brood removal. To test this, we treated patches of ~30 cells with β-ocimene or oleic acid, and separated the patches by either one band of untreated cells ('near' treatments) or located the patches on two different frames, with two untreated brood frames separating them ('far' treatments). We did this for N = 5 colonies, and found no effect of patch proximity on brood removal rates (Figure 4.1 C; two-way ANOVA; factors: odorant, proximity; $F = 0.025$, $p = 0.88$). In another test, we found that

workers removed treated pupae and prepupae at similar rates (Figure 4.1 D; four-factor ANOVA; factors: dose, odorant, hygienicity, age; $F = 0.84$; $p = 0.36$; see Table 4.2 for sample sizes). Therefore, we combined data for the two ages and used the front-way assay to test if colonies with higher hygienicity responded to the odorants differently than colonies with lower hygienicity.

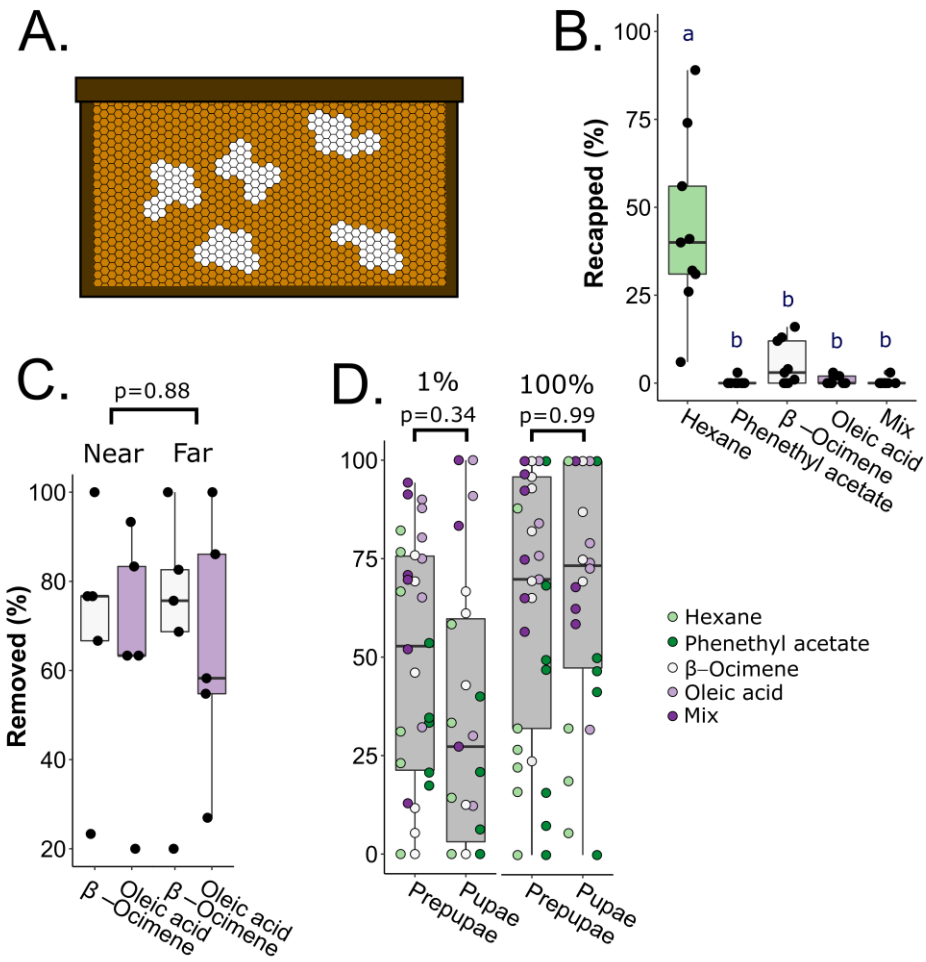


Figure 4.1. Front-way odorant assay preliminary tests.

A. Schematic of the front-way assay. Patches of capped brood (~30 cells in technical duplicate per colony) developing naturally in a standard frame were uncapped (white patches) and 1 µl of odorants (β-ocimene, oleic acid, a 1:1 v/v mix of the two, phenethyl acetate or hexane) at either 1% or 100% concentrations (v/v in hexane) were dispensed onto the brood. Frames were incubated in the colony's brood box for 3 hours before recording removal rates. B. Post-front-way assay recapping frequencies. Data from N = 9 colonies were analyzed with a one-way ANOVA (level: odorant; F = 13.3, p = 2.4e-8) followed by a Tukey HSD test. Letters indicate groups that are significantly different from one another Tukey HSD p < 0.05). C. Preliminary test for a patch proximity effect. N = 5 colonies were tested, varying the distance between β-ocimene and oleic acid patches (near = patches on the same frame, separated by one band of untreated capped brood; far = patches on different frames separated by two untreated brood frames). We analyzed the data by a two-way ANOVA and found no effect of patch (F = 0.025, p = 0.88) nor interactive effect between patch and odorant (F = 0, p = 1.0). D. Preliminary test for a brood age effect. We performed the front-way assay on N = 9 colonies and calculated the percent prepupa and pupa removal. Due to variability in patch composition, not every colony had the same number of replicates for each stage and dose (see Table 4.2 for all sample sizes). Data were analyzed with a four-way ANOVA (levels: odorant, age, hygienicity, dose), which identified no significant effect of age nor interactions with any other factors, followed by a Tukey HSD test. 1% and 100% refer to odorant concentrations. All boxes depict the interquartile range (IQR) and the whiskers span 1.5*IQR.

Table 4.2. Replicate information for age-related brood removal measurements

	Prepupae		Pupae	
	1%	100%	1%	100%
Hexane	6	6	6	4
β-ocimene	6	7	5	6
Oleic acid	6	6	4	5
Blend	6	7	3	5
Phenethyl acetate	6	7	5	6

4.3.2 **Effect of hygienicity**

We tested N = 5 colonies with high hygienicity (freeze-killed brood score > 80%) and N = 5 colonies with low hygienicity (freeze-killed brood score < 80%) (Figure 4.2 A), and found significant effects of dose, odorant, and hygienicity (Figure 4.2 B; three-factor ANOVA; dose: F = 61.2, p = 4.3e-11; odorant: F = 19.8; p = 7.1e-11; hygienicity: F = 20.2, p = 2.7e-5). Normally, only colonies with ≥95% FKB assay scores are considered "highly hygienic;" however, we lowered our threshold here to 80% in order to achieve a balanced experimental design. As expected, brood treated with neat odorants were removed significantly more frequently compared to those treated with diluted odorants. We had intended phenethyl acetate to be a positive control odorant, but surprisingly, we found that it induced similar brood removal as the negative control (hexane), both of which were the lowest of all those we tested. In the neat odorant treatments, β-ocimene, oleic acid and their blend all induced significantly higher brood removal relative to hexane (Tukey HSD; p = 0.0034, p = 0.0075, and p = 0.0049 respectively), but in the diluted odorant treatments, none of the odorants induced significantly different brood removal. However, their relative patterns still reflect what's observed in the neat odorant treatments.

We expected colonies with higher hygienicity to respond more strongly to the odorant stimuli than colonies with lower hygienicity. We found that indeed, the higher hygienicity colonies removed significantly more treated brood overall in both the neat odorant treatments (Tukey HSD; p = 0.0084), as well as the diluted treatments (p = 0.011). This agrees with previous electroantennography studies showing that hygienic bees' antennae are more sensitive to disease odorants than non-hygienic bees[91,92].

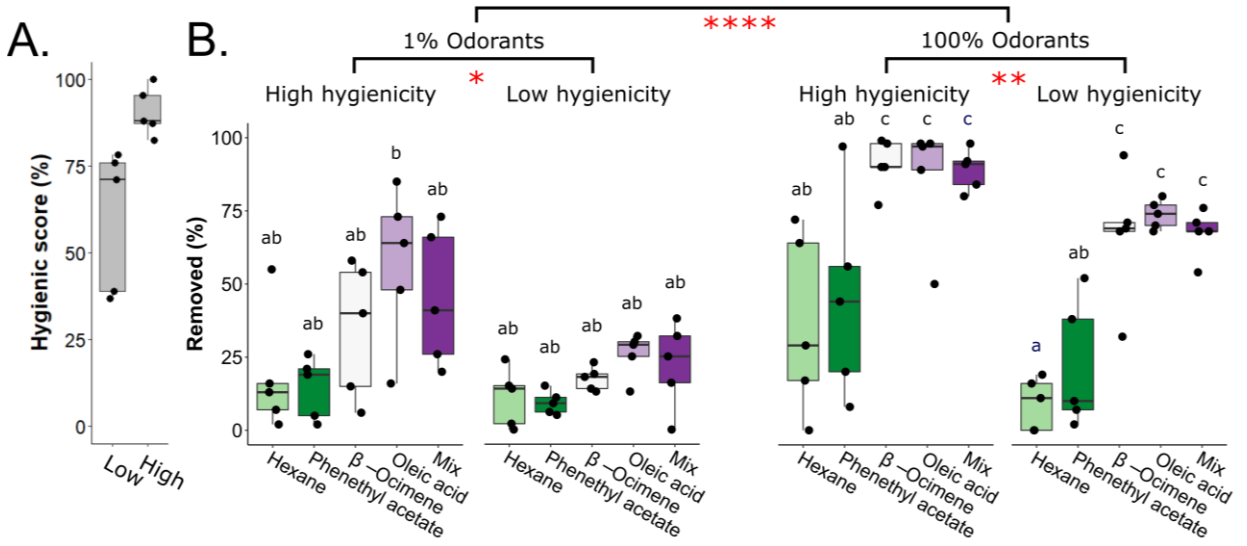


Figure 4.2. Front-way odorant assays to investigate effects of hygienicity.

A. Distribution of hygienic scores for the tested colonies. 10 colonies were tested in total. The lowest-scoring 5 were assigned to the 'low hygienicity' group (scores < 80%) and the highest-scoring 5 were assigned to the 'high hygienicity' group (scores > 80%). B. Post-front-way assay removal frequencies. Hexane is the negative control and phenethyl acetate (a chalkbrood odorant) was meant to be the positive control. Data from 5 low hygienicity and 5 high hygienicity hives were analyzed with a three-factor ANOVA (levels: dose, odorant, hygienicity; dose: F = 61.2, p = 4.3e-11; odorant: F = 19.8; p = 7.1e-11; hygienicity: F = 20.2, p = 2.7e-5), followed by a Tukey HSD post-hoc test. Significance code (Tukey HSD): * p < 0.05, ** p < 0.01, **** p < 0.0001. Boxes depict the interquartile range (IQR) and the whiskers span 1.5*IQR. Letters indicate groups that are significantly different from one another at Tukey HSD p < 0.05).

4.4 Back-way odorant assay

The front-way odorant assay is a quick method of gauging if odorants can induce brood removal, but it cannot test for odorant transmission through the physical barrier of the wax cap. To investigate the odorants in a more realistic scenario, we used the Jenter™ system that allows us to treat brood with odorants while maintaining the integrity of the brood cells, as first described by Boecking and Drescher[231]. We call this the back-way odorant assay (Figure 4.3 A), since we add the odorants through the back of the brood cell. Briefly, we place a queen in a Jenter™ cage until she lays eggs in the comb of the cage, then release her and allow the workers to rear the brood until it is capped. The back of the Jenter™ cage is equipped with removable plugs that enable odorants to be added inside the cell without disturbing the delicate wax cell cap, and plugged again to close the brood cell. We used this method to add neat hexane, β-ocimene, oleic acid and the odorant blend to 9-10 brood cells each, before and after pupation (N = 5 colonies for each age). We found that after incubating in the hive for 20 h, β-ocimene did not induce significantly more brood removal relative to hexane (Figure 4.3 B; two-factor ANOVA followed by Tukey HSD; $p = 0.82$ for pre-pupal brood and $p = 0.10$ for post-pupal brood). However, oleic acid strongly induced pre-pupal removal ($p = 0.0004$) and marginally non-significant post-pupal removal ($p = 0.057$). The odorant blend induced the most consistently high brood removal of them all, which was significant for both brood ages ($p = 0.0004$ for pre-pupal and $p = 0.0003$ for post-pupal).

The next year, we conducted similar back-way odorant assays using diluted (1%) odorants instead of neat odorants. Only pupae were analyzed (N = 4 colonies). We found the same general

trend as for the pupae treated with neat odorants, although there were fewer brood removed overall, as expected (Figure 4.4). Hexane still induced low levels of brood removal, and the mixed odorants induced the highest, although these differences were not significant.

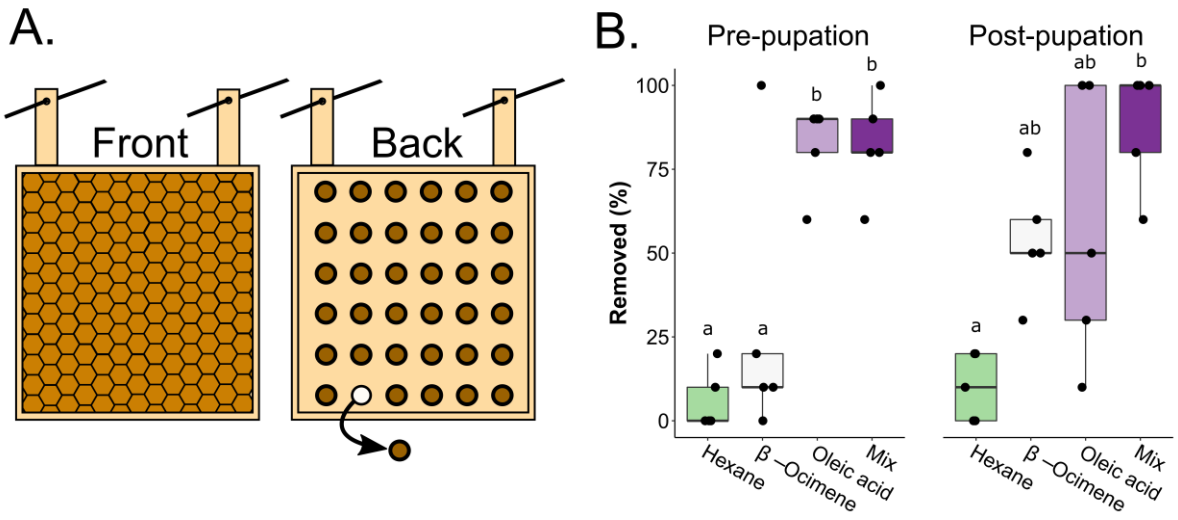


Figure 4.3. Back-way odorant addition assays with neat odorants.

A. Schematic of the back-way assay. Queens were caged in a Jenter™ queen rearing cage (a hanging square of artificial comb) until she populated the cells with eggs. The queens were released and brood were allowed to develop until capping (front view). We treated brood cells with neat odorants in a semi-random design through the cell plugs (back view, brown circles), then the odorant-impregnated brood was incubated in the colony for 20 h to allow time for odorant diffusion, uncapping, and removal. Diagrams are not to scale. The actual Jenter™ cage has ~ 100 removable plugs (one every 3rd cell). B. We treated pre-pupal and post-pupal brood with each odorant (9-10 brood cells for each age and odorant, N = 5 colonies). Data was analyzed using a two-factor ANOVA (levels: age and odorant) followed by a Tukey HSD post hoc test. There was a significant effect of odorant (F = 20.3, p = 1.51e-7), no significant effect of age (F = 0.16, p = 0.694), and no significant interactive effect (F = 1.9, p = 0.157). Letters indicate groups that are significantly different from one another (Tukey HSD p < 0.05). Boxes depict the interquartile range (IQR) and whiskers span 1.5*IQR.

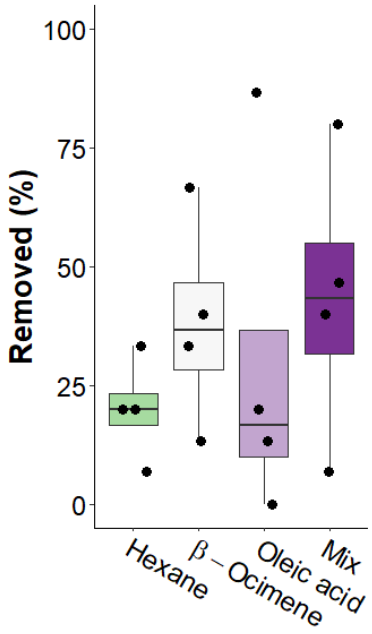


Figure 4.4. Backway odorant addition assay with diluted odorants.

The backway odorant addition assay was performed on white-bodied pupae from N = 4 colonies as described for Figure 4.3, except diluted (1% in hexane) odorants were utilized instead of neat odorants and patches were composed of 13-15 pupae. No significant differences were identified (one-way ANOVA). Boxes depict the interquartile range (IQR) and whiskers span 1.5*IQR.

4.5 Characterizing the age-dependent background odorant profile

To try to explain the patterns of pre-pupation and post-pupation brood removal, we investigated changes in the background volatile and non-volatile odorant profiles that could confound with our odorant treatments. To do this, we performed solid-phase micro-extraction gas chromatography-mass spectrometry (SPME-GC-MS) on extracts from 5th instar larvae, prepupae, and pupae. We analyzed N = 5 independent brood, from 5 different colonies, for each stage. We also used a hexane wash (with the same replicate structure as before) to extract cuticle compounds from these life stages and analyzed them by GC-MS as well, capturing the less volatile signals. We found that β-ocimene abundance changed most significantly according to age (one-way ANOVA, Benjamini-Hochberg corrected 1% FDR; $p = 0.0010$, $q = 0.01$), with

relatively high amounts emitted in 5th instar larvae and prepupae, and low amounts in pupae (Figure 4.4). Two other minor chromatogram components were also differentially emitted (compounds 2 and 4, corresponding to isopropanol and 2-pentanone, respectively). Other volatile compound identifications are reported in Table 4.3. The hexane wash identified many branched chain hydrocarbons which were differentially emitted with age but importantly, oleic acid was not among the identified molecules for any of the three developmental stages (Table 4.4).

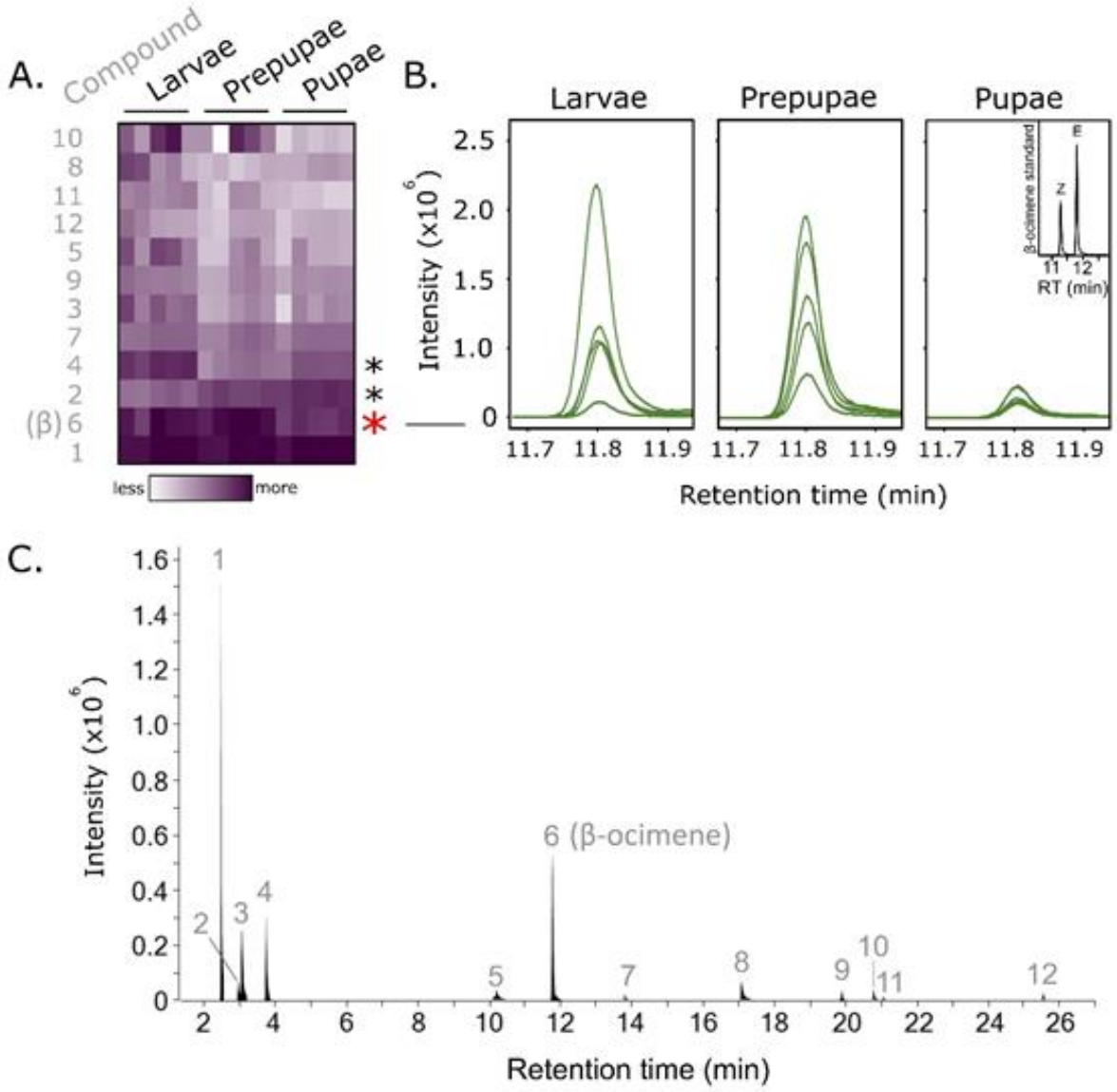


Figure 4.5. β-ocimene abundance in larvae, prepupae and pupae.

We performed solid phase micro-extraction gas chromatography mass spectrometry (SPME-GC-MS) on extracts from 5th instar larvae, prepupae and pupae (N = 5 colonies each). A. Heatmap showing intensities of all integrated peaks. Areas under the curve were compared between ages using a one-way ANOVA and Benjamini-Hochberg correction (5% FDR). Each row corresponds to peak intensities belonging to a different compound. β-ocimene, the most significantly different compound, is indicated with a red asterisk, while two other significantly different compounds (matching to isopropanol (2) and 2-pentanone (4)) are indicated with black asterisks. Raw GC-MS data is available at http://github.com/AlisonMcAfee/test. B. Chromatogram traces of the β-ocimene peak. Its identity was confirmed with a synthetic standard (inset chromatogram). Based on its retention time, only the E isomer was identified in the brood. C. Example SPME-GC-MS total ion chromatogram. Numbers correspond to compounds labelled in A. Further compound identity and abundance information is available in Table 4.3.

Table 4.3. Log10 transformed SPME-GC-MS peak areas for N = 5 brood from each stage from 5 different colonies

			Larvae		Prepupae		Pupae	
***Peak #**	**Retention time (min)**	**Proposed identity**	**Ave**	**St. Dev.**	**Ave**	**St. Dev.**	**Ave**	**St. Dev.**
1	2.49	Thiobis-methane	6.50	0.13	6.48	0.14	6.53	0.12
2	2.97	2-propanone	4.89	0.16	5.49	0.10	5.81	0.12
3	3.04	Octane	4.99	0.33	4.07	0.27	4.08	0.53
4	3.75	Isopropanol	5.81	0.16	4.61	0.23	5.11	0.25
5	10.23	Isoamyl acetoacetate	4.98	0.50	3.93	0.50	3.82	0.36
6**	11.78	β-cis-Ocimene	6.22	0.39	6.43	0.23	5.73	0.15
7	13.81	β-octahydroindoloquinolizine	4.88	0.10	4.69	0.22	4.77	0.21
8	17.17	Acetic acid	4.66	0.67	3.58	0.19	3.96	0.16
9	19.94	Ethyl 4-(chloromethylene)-2,2-diphenyl-3-oxazoline-5-carboxylate	4.63	0.14	4.17	0.30	4.20	0.18
10	20.78	Propanoic acid	5.09	0.87	4.50	1.29	3.43	0.22
11	21.07	n/a	4.48	0.11	3.80	0.44	3.36	0.14
12	25.56	1,3-diphenyl-1-trimethylsilyloxy-1-pentene	4.20	0.36	3.75	0.29	3.68	0.18

*Only peaks with > 4,000 cts apex intensity are displayed

**Compound identity confirmed as β-ocimene based on spectral matching and comparing the retention time to a standard.

Ave = Average

St. Dev. = Standard deviation

Table 4.4. Log10 transformed hexane wash GC-MS peak areas for N = 5 brood from each stage from 5 different colonies

			Larvae		Prepupae		Pupae	
***Peak #**	**RT**	**Proposed identity**	**Ave.**	**St. Dev.**	**Ave.**	**St. Dev.**	**Ave.**	**St. Dev.**
1	14.78	Triacontane	6.24	0.21	6.02	0.10	6.46	0.07
2	14.93	2-methyl hexadecane	5.83	0.09	6.18	0.12	6.27	0.06
3	15.91	n-pentacosane	6.78	0.19	6.57	0.12	6.92	0.02
4	16.07	Heptadecane	6.41	0.11	6.90	0.10	6.84	0.04
5	17.27	Octacosane	7.04	0.09	7.29	0.09	7.64	0.02
6	17.47	1-eicosanol	6.82	0.11	7.47	0.09	7.56	0.03
7	18.09	Tricosane	5.44	0.10	6.08	0.09	6.58	0.02
8	18.31	Hexatriacontane	5.35	0.18	6.29	0.09	6.58	0.02
9	19.14	Nonacosane	6.50	0.06	6.83	0.12	7.42	0.03
10	19.41	2-methyl-octadecane	6.55	0.23	7.15	0.08	7.56	0.01
11	22.01	Tetracosane	6.07	0.07	6.13	0.10	6.67	0.07
12	22.41	2-methyl-eicosane	6.56	0.28	6.82	0.08	7.31	0.02
13	27.28	2-methyl-octadecane	6.24	0.27	6.42	0.14	6.84	0.05

*Only peaks with > 4,000 cts apex intensity are displayed
Ave = Average
St. Dev. = Standard deviation

4.6 Broodless back-way odorant assay

In our previous odorant addition assays, it is possible that the removal apparently induced by the odorants is actually a result of the brood's response to the topical application. Based on our toxicity assays, we cannot rule out this potential effect. Therefore, we performed broodless odorant addition assays using the same back-way addition described earlier, but we replaced the brood with 1 µl of either neat or diluted odorant-impregnated strips of filter paper (Figure 4.6). Surprisingly, cells containing filter papers treated with hexane were removed at the highest frequency, when before, hexane-treated live brood was consistently removed with the lowest frequency (Figure 4.3). This suggests that there is also a "live" signal emitted by healthy brood which suppresses hygienic behaviour. Contrary to what we had expected, cells containing filter

paper treated with 100% β-ocimene or a 100% blend (1:1) of β-ocimene and oleic acid were removed least often. Cells containing oleic acid alone were removed at rates comparable to hexane and empty cells. Therefore, in this context (*i.e.* against a background of no live brood odorants at all), β-ocimene appears to actually suppress hygienic behaviour, serving more like a live signal than a death signal. This suggests that the context in which worker bees encounter odorants is an extremely important factor dictating their decision-making and ultimately their behavioural output.

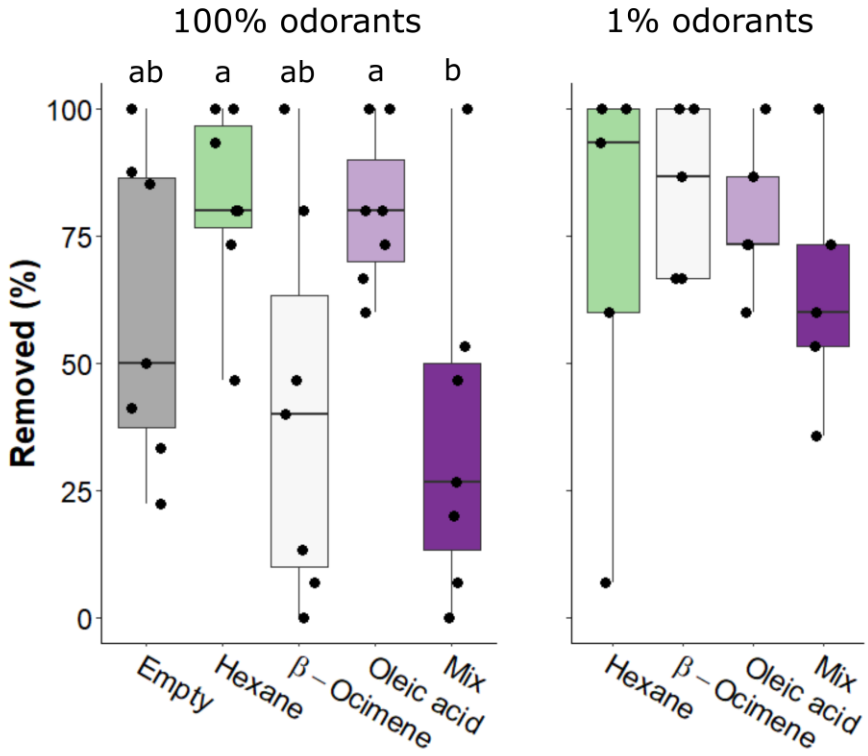


Figure 4.6. Broodless back-way odorant addition assays.

Broodless back-way odorant addition assays were performed as described in Figure 4.3, but instead of applying odorants to live pupae, the brood was removed through the back of the Jenter set and replaced with odorant-impregnated strips of filter paper (1 µl of either neat or diluted odorants). Semi-random patches of 14-15 pupae were removed for each odorant. N = 6 colonies were analyzed for neat odorants and N = 5 colonies were analyzed for diluted odorants. The responses to neat odorants were analyzed with a one-way ANOVA. There was a significant effect of odorant ($F = 3.83$, $p = 0.0125$), and letters indicate groups that were significantly different following a Tukey HSD test ($p \leq 0.05$). We did not perform statistical tests on the responses to diluted odorants. Boxes depict the interquartile range (IQR) and whiskers span 1.5*IQR.

Although β-ocimene appears to suppress hygienic behaviour in the broodless back-way odorant assays, there are likely other "live" signals contributing to normal hygienic behaviour suppression. To attempt to identify what these signals may be, we performed another broodless back-way assay with two different extracts from live brood: one being a hexane cuticle wash of prepupae, and another being prepupae crushed in hexane (both extracts used 15 prepupae in 500 µl hexane). None of the extracts suppressed hygienic brood removal compared to either hexane alone or empty brood cells (Figure 4.7).

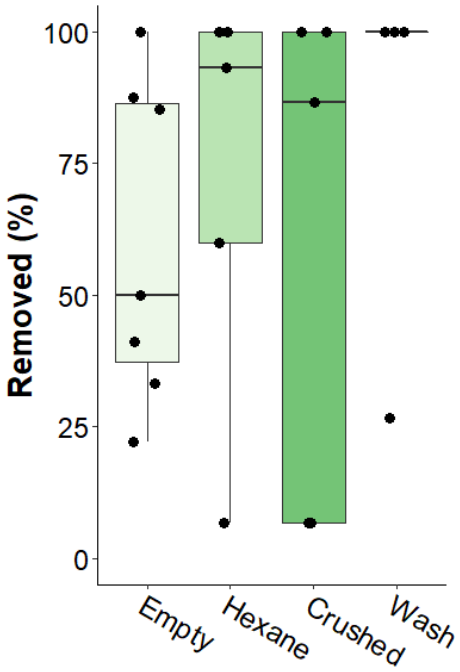


Figure 4.7. Broodless back-way odorant assays using extracts from live brood.

The Broodless back-way odorant assay was performed as described in Figure 4.6. Empty treatments and hexane treatments are the same as in Figure 4.6 for comparison. Crushed refers to an extract of compounds obtained from crushing 15 prepupae in 500 µl hexane with a glass rod, then letting the phases separate for 10 minutes, followed by extracting the supernatant from the vial with a syringe. The wash treatment refers to an extract obtained by washing the cuticles of 15 prepupae with 500 µl hexane in a glass vial, and extracting the wash solution with a syringe. 1 µl of all substances was applied to strips of filter paper for the assay.

4.7 Electroantennography at hive realistic temperatures

Previously, we reported that stimulating honey bee antennae with oleic acid yielded no measurable nerve depolarization signal above the background stimulus of air alone[230]. Since we observe that in some cases oleic acid can induce hygienic behaviour in brood removal assays (including when the brood cell cap remains in-tact), we questioned if the workers were detecting oleic acid-treated cells by olfaction or some other sense (*e.g.* gustation). To investigate this further, we replicated the electroantennography experiment (N = 13 left antennae and N = 14 right antennae) comparing oleic acid to background stimulation, but at a temperature that better-matches in-hive conditions. When we administered warmed oleic acid (at approximately 33ºC), we found that it stimulates worker antennae only slightly more than blank stimuli (Figure 4.8). There was also a significant effect of odorant (two-way ANOVA; levels: odorant, side; F = 12.4; p = 2.3e-5), with β-ocimene and the odorant blend inducing significantly higher antennal nerve depolarizations than oleic acid in left antennae (p = 0.011 and p = 0.016, respectively). The same comparisons yielded a marginally non-significant response in the right antennae (p = 0.085 and p = 0.086, respectively).

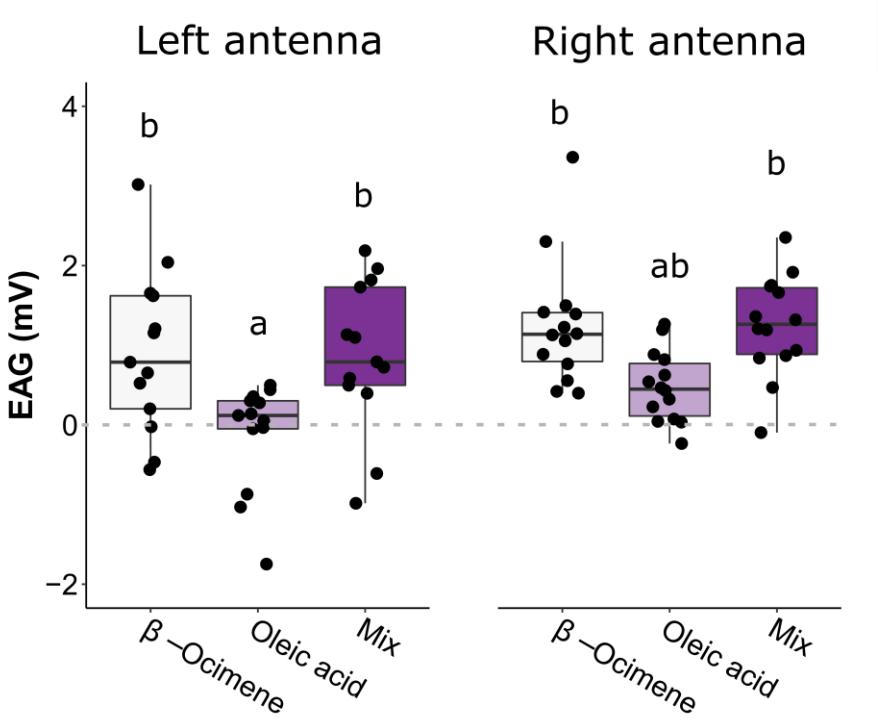


Figure 4.8. Electroantennography (EAG) responses to odorants.

We excised left (N = 13) and right (N = 14) antennae from honey bees in a single highly hygienic colony (score = 95%) and measured the EAG response to neat odorants (Syntech™ CS-55) at hive-realistic temperatures (around 33ºC). The EAG response represents blank-subtracted odorant stimuli. We found a significant effect of odorant (two-way ANOVA; levels: side, odorant; F = 12.4, p = 2.3e-5), and letters indicate groups that are significantly different from one another (Tukey HSD p < 0.05). Boxes depict the interquartile range (IQR) and whiskers span 1.5*IQR.

4.8 Conclusions and future directions

In the present work, we investigate two candidate hygienic behaviour inducers that are emitted from freeze-killed brood – β-ocimene (a co-opted pheromone emitted by brood and queens[222,232-235]) and oleic acid (a well-known necromone and necrophobic compound in other arthropods[66,69,70,78,97-102]) – using *in vivo* and *ex vivo* techniques. We demonstrated 1) that treating brood with the odorants is sufficient to induce hygienic behaviour in bioassays (Figure 4.2 and 4.3), 2) hygienic behaviour is normally suppressed by a yet-unidentified "live" signal emitted

from healthy brood (Figure 4.6 and 4.7**)**, and 3) despite being a non-volatile compound, oleic acid can stimulate nerve depolarizations worker antennae at hive temperatures, although the magnitude of stimulation is very low (Figure 4.8). Although β-ocimene and oleic acid were identified from freeze-killed brood and do not extrapolate to all brood diseases, it is a relevant model with which to investigate some molecular interactions governing this complex behavioural process. This is not the first time that a brood pheromone has been implicated in social immunity; Mondet *et al*.[96] found that *Varroa*-infested brood produced elevated levels of brood ester pheromone. Other researchers have found that oleic acid is both contained and emitted by *Varroa,* on top of it being generally associated with insect death[203,236-238].

4.8.1 **Potential cooperation between β-ocimene and oleic acid**

β-ocimene and oleic acid have very different chemical properties: β-ocimene is a volatile alkene (boiling point: 65-66^{o}C) and oleic acid is a viscous, mono-unsaturated carboxylic acid (boiling point: 360^{o}C). Both are emitted more strongly from freeze-killed honey bee brood compared to live brood[230], but based on their differences in volatility, we expect them to permeate the brood cell cap at different rates. In a biologically relevant scenario, this spatial diffusion should be necessary for adult workers to detect odorant signals evolving under the cap. Since the odorant blend induces the highest brood removal most consistently in the back-way odorant assays (Figure 4.3 and Figure 4.4), but not the front-way odorant assays (Figure 4.2), we suggest that β-ocimene and oleic acid may be acting in a cooperative manner when there is a wax cap barrier in place. Since our electroantennogram recordings (Figure 3.8) show that there is no synergistic effect at the level of antennal odorant detection, we suggest they could instead be cooperating via volatility mechanics. For example, a potential mechanism is that β-ocimene diffuses rapidly and

attracts worker visits (as it is already known to do for larval feeding[233]) and after subsequent cell inspection, oleic acid acts as the determinant death cue that stimulates brood removal. In the front-way odorant assay, however, the workers are in constant contact with the odorants (since there is no cap acting as a barrier); therefore, oleic acid is readily detectable even in the absence of an attractant.

The back-way odorant assay we describe here is the most biologically relevant assay employed for testing different odorants' abilities to induce hygienic behaviour. Unlike other behavioural assays where cells are either uncapped (as in our front-way odorant assay) or filled with odorant-impregnated brood dummies[94,95], this assay fully maintains comb integrity and allows the workers to perform the complete behaviour (uncapping and removal). While the odorant blend was the most consistently high inducer of brood removal, oleic acid alone also induced significant brood removal for young (pre-pupal) brood, but not post-pupal brood (Figure 4.3 B). Based on our analysis of the background brood odorant profile, this could be because of naturally released β-ocimene (Figure 4.5) interfering with the synthetic odorant treatments. Since the younger brood emitted significantly more natural β-ocimene compared to the older brood, the young brood treated with oleic acid was, in a way, also a blend, which could in part explain why this treatment induced similar removal to the synthetic blend for the pre-pupal brood but not post-pupal. Very few pre-pupal β-ocimene-treated brood were removed (28%), which is consistent with young brood emitting their own β-ocimene already. Post-pupal β-ocimene-treated brood, which emit very little natural β-ocimene, were removed at higher rates (54%), although this was not statistically significant ($p = 0.10$). More replicates are necessary to determine if this is a consistent trend, but it could be that β-ocimene serves as a hygienic behaviour inducer only

when significantly more of this compound is emitted than expected based on the emanating brood's age (or age of the brood immediately surrounding it). This is consistent with the proposal by Mondet *et al*.[96] that developmental delay may be a signal targeting brood for removal by VSH bees.

In the broodless odorant assays, we found that the cells containing filter papers treated with the negative control odorant (hexane) were actually targeted for the most uncapping behaviour. Empty cells were targeted at a similar frequency, despite containing no odorant and no filter paper. This rules out the possibility that some compounds from the filter paper itself were causing hygienic behaviour. Rather, there must be signals coming from live brood which suppress the behaviour under normal conditions, but we were unable to identify a live brood solvent extract which contains them. β-ocimene and the blend of β-ocimene and oleic acid appear to have a suppressive effect, suggesting that the behaviour(s) induced by β-ocimene can vary greatly depending on the context. Given that this compound has many known biological functions in the hive already, perhaps this should not be surprising. Overall, the evidence suggests that the context in which odorants are encountered is very important for dictating the resulting behaviour.

In Chapter 3, we identified carbon dioxide as the only compound that was emitted more strongly from live brood compared to dead brood. Since honey bees (and other social insects) can detect carbon dioxide,[239] it is possible that the emission of carbon dioxide with active respiration is one brood signal that suppresses hygienic behaviour under normal conditions. This hypothesis has yet to be tested.

4.8.2 **Inconsistencies regarding phenethyl acetate**

Swanson *et al.*[94] originally identified phenethyl acetate as a strong hygienic behaviour-inducing compound emitted from chalkbrood-infected larvae; however, in our experiments, we found that it induces similar levels of hygienic behaviour relative to the negative control in both the diluted ($p = 0.99$) and neat ($p = 0.97$) odorant treatments, which is less than both oleic acid and β-ocimene. In fact, Swanson *et al.* found that phenethyl acetate induced 40-100% brood removal using 50% of the odorant amount we used. One reason why we did not observe high phenethyl acetate removal rates could simply be because we did not analyze many highly hygienic colonies (≥95% FKB score). However, it could also be that the colonies used by Swanson *et al.*[94] were from a genetic lineage that was more sensitive to chalkbrood odorants than ours. Indeed, the two populations of colonies are geographically isolated and are likely adapted to different climates, conditions, and disease challenges. Furthermore, the surprisingly low degree of overlap between differential expression studies comparing hygienic and non-hygienic bees suggests that there are many adaptive routes for bees to become hygienic[111]. It could simply be that the hygienic bees used in experiments by Swanson *et al.*[94] possess different molecular machinery that allows them to be sensitive to different disease odorants than the colonies used in the present study.

4.8.3 **Caveats to the behavioural assays**

Based on our data, we cannot yet rule out the possibility that some of the behavioural response toward odorant-treated brood was a result of toxicity of the odorant itself. In an acute toxicity assay we found that odorants could cause contact toxicity when dispensed on the abdomen of pupae, inducing a prophenoloxidase immune response (Table 4.1). However, 100% of hexane-

treated brood developed normally. Removal rates for the front- and back-way odorant addition assays do not appear to mirror the results of the toxicity assay, so it is unlikely that the brood removal was entirely due to contact toxicity. That being said, the only toxicity outcomes we measured was the prophenoloxidase response. There could be other cues that odorant contact stimulates the brood to emit, which we did not measure. In addition, we only investigated abdominal contact toxicity, which is the application site for the back-way assays, whereas in the front-way assays, we applied odorants to the head, which could yield a different response. Other limitations include that the toxicity outcome was measured after a period of 1 d or 2.5 d, whereas other developmental effects could take longer to appear. We note, however, that 2.5 d is much longer than the duration of any of our behavioural assays here, and the results of separate toxicity assays after 1 d and 2.5 d are not consistent. In addition, we tested only pupae in the toxicity assay, and not 5^{th} instar larvae or prepupae, which could respond differently to the odorants. These are all important caveats to this work, and warrant further investigation.

One way these concerns can be addressed in the future is by developing an assay utilizing brood dummies instead of real brood to eliminate the brood effect. Swanson *et al.*[27] developed a similar assay using brood ester pheromone- and odorant-impregnated paraffin brood dummies in open cells, measuring cell capping (non-hygienic activity) and capping refrainment (hygienic activity) as a proxy for hygienic behaviour, since worker bees cannot physically remove the paraffin brood dummies from the cells. This eliminates the brood effect, but has the caveat that leaving a cell uncapped is not the same as performing hygienic behaviour. In our front-way experiments, we noticed that cells were frequently left both uncapped and uncannibalized – an outcome which would count as hygienic activity if using paraffin brood dummies. To get around this, we

developed a broodless hygienic test that still allows the object to be removed. We achieved this by removing developing brood through the back of a Jenter™ set and replacing it with a small odorant-treated piece of filter paper but, as described earlier, this has raised more questions than it has answered.

On one hand, our 100% odorant treatments (1 µl) could be criticized as not being biologically relevant because the signal is too strong; however, this may work to our advantage to overcome the brood effect. By using such a strong odorant signal in the front-way assays, and measuring the behaviour response after a short period of time (3 h, compared to 24 h for the standard freeze-killed brood assay to measure hygienicity), this should a) minimize the amount of time the brood has to produce a strong response, and b) the experimental treatment should be the dominant signal. For the back-way assays, a longer incubation period (20 h) was utilized since in preliminary tests the behavioural response after 3 h was too low to be useful. This means that there was more time for a potential brood effect to evolve, which may have impacted our results.

4.8.4 **Conclusion**

In summary, this data suggests that oleic acid and β-ocimene induce brood removal in honey bees. Bees with higher hygienicity respond to the odorants more strongly than bees with lower hygienicity, and the blend induces brood removal most consistently in the most biologically realistic brood removal assay. Despite being non-volatile, oleic acid appears to be detectable even beneath a brood cell cap; however, it is possible that the bees are detecting the brood's reaction to the odorant rather than the odorant alone. Our electrophysiology tests show that oleic acid only marginally stimulates antennal nerve responses in environmental conditions similar to

those inside a hive, suggesting that if they are detecting the odorant alone, extremely close proximity would be necessary for bees to detect it. Oleic acid elicits necrophoretic and necrophobic behaviour across phylum Arthropoda[66,69,70,78,97-102], and these data piece its activity in honey bees into the phylogenetic puzzle. To the best of our knowledge, our data shows for the first time that this 'death cue' function is evolutionarily conserved in honey bees, and that oleic acid may be working in concert with β-ocimene as an attractant. Future experiments will be necessary to eliminate the possibility of an odorant-induced brood effect contributing to these results, and to confidently identify the suppressive "live" cue suggested by the broodless odorant assays.

4.9 Methods

4.9.1 Data access

The raw GC-MS data is available for download at http://www.github.com/AlisonMcAfee/test.

4.9.2 Honey bee colonies and hygienic testing

All hygienic testing, sampling and odorant assays were completed during the summers of 2017 with the exception of the broodless back-way odorant assays and 1% back-way assays with brood, which were conducted in the summer of 2018. Hygienic testing was performed as described in Section 3.8.3.

4.9.3 Front-way odorant assays

To perform the front-way odorant assays, we retrieved two brood frames from each colony, uncapped patches of brood with tweezers and dispensed 1 µl of odorant treatments onto the

exposed brood (Figure 4.1 A). Wax caps were not replaced after odorant addition. We tested the odorants β-ocimene, oleic acid, a 1:1 v/v blend of the two, phenethyl acetate (positive control), and hexane (negative control) at concentrations of 100% and 1% (v/v in hexane). Phenethyl acetate was not included in the blend because it is not known to co-occur with the other odorants (phenethyl acetate is from chalkbrood, while β-ocimene and oleic acid are associated with freeze-killed brood). For each odorant and concentration, we performed two technical replicates (2 patches of 30 brood cells each, one on each frame). We tested the different concentrations on different days. After treating the brood patches with odorants, we photographed, traced, and labelled each patch on a transparency and replaced the brood frame in the hive. After 3 h, we returned to the hive and recorded the number of brood cells that were cannibalized and partially cannibalized (cumulatively yielding the number 'removed') or recapped.

Brood patches were composed of variable developmental stages (mostly prepupae and pupae, but some 5^{th} instar larvae; Table 4.5), so we used the photographs from pre- and post-incubation to assess the fraction of each developmental stage that were removed and/or recapped by the workers. With a clear anterior view, the prepupae can be distinguished from 5^{th} instar larvae based on their upright, elongated body and a 'crook-neck' appearance. Due to variable patch composition, we did not obtain the same number of biological replicates for every developmental stage and odorant (see Table 4.2 for complete replicate information for each stage and odorant concentration). Data for 5^{th} instar larvae are not shown because too few patches contained them to reliably test if there was a differential response to larvae (they made up < 10% of tested brood cells overall). This is because the time between cell capping and transforming to a prepupa is very short – in the order of hours – so catching this stage in a naturally laid comb is infrequent.

These sparse data were therefore excluded from subsequent analyses. In a preliminary test, brood removal data were analyzed with a four-factor ANOVA (levels: dose, odorant, age, hygienicity) followed by a Tukey HSD to determine if there was an effect of age between prepupae and pupae. Since there was no significant effect of age alone ($F = 0.87$; $p = 0.36$) nor in combination with any other factors (odorant*age: $p = 0.61$, dose*age: $p = 0.15$, hygienicity*age: $p = 0.79$, odorant*dose*age: $p = 0.58$, odorant*hygienicity*age: $p = 0.73$, dose*hygienicity*age: $p = 0.17$, odorant*dose*hygienicity*age: $p = 0.71$), we pooled the pupa and prepupa data for subsequent analyses. All statistical analyses were performed in R unless otherwise specified.

Table 4.5. Frequencies of brood ages across all front-way odour assays.

	Total	% Larvae	% Prepupae	% Pupae
Hexane	822	10.5	60.1	29.4
Mix	853	12.5	62.1	25.3
Ocimene	817	6.2	63.8	30.0
Oleic acid	818	8.8	65.0	26.2
Phenethyl acetate	803	8.5	62.5	29.0
Overall	4113	9.3	62.7	28.0

In a second preliminary experiment, we confirmed that there was no effect of patch proximity in the front-way odorant assay. We varied proximity by testing two patches of brood per colony that were either separated by a single capped cell-width on the same side of a frame (‘near’), or on different frames with two brood frames located between them (‘far,’ $N = 5$ colonies each). One microliter of oleic acid (the least volatile odorant tested) or β-ocimene (the most volatile odorant tested) was added to the cells of each patch. The data were analyzed with a two-factor ANOVA (levels: proximity, odorant).

To assess the relationship between hygienicity and odorant-treated brood removal, we performed the front-way odorant assay on 10 colonies (and two technical replicates per colony, which were averaged to produce one biological replicate) with varying hygienic score (39% to 100%). We grouped the colonies into N = 5 with higher hygienicity (scoring > 80%), and N = 5 with lower hygienicity (scoring < 80%) (Figure 4.2 A). As before, we removed the larval cells from the analysis (~10% overall) and since we previously determined that there was no effect of brood age between prepupae and pupae in the front-way assay, we did not distinguish between these stages statistically. These data were analyzed using a three-factor ANOVA (levels: dose, odorant, hygienicity) followed by a Tukey HSD post hoc test. Brood recapping data was derived from the same assays (N = 9 for each odorant (data was unavailable for one colony)) using a one-way ANOVA (level: odorant).

4.9.4 **Back-way odorant assays**

To test the effects of β-ocimene, oleic acid and their 1:1 v/v blend in a more biologically realistic scenario, we developed the back-way odorant assay (Figure 4.3 A). This assay adapts artificial comb cages of the Jenter™ queen rearing system to instead rear worker brood *in situ*. The Jenter™ set features removable plastic plugs from the rear of the comb – usually used to harvest eggs/larvae for queen rearing – which provide convenient access points for odorant addition without damaging the wax brood cell caps or the brood itself.

We conditioned the Jenter™ comb cages by placing them in a colony for several days, allowing the bees to draw out full-height comb cells. We then caged the queens and allowed them sufficient time to populate the combs with eggs (typically overnight). We released the queens and allowed the workers to rear the brood *in situ*. Once capped, we inspected the brood via the removable plugs to confirm the developmental stage. Through this small posterior window, 5th instar larvae and prepupae are indistinguishable, but pupae are easily recognized by their clearly developed abdomen and hind tarsi. This is in contrast to the front-way odorant assay, where 5th instar larvae and prepupae are distinguishable due to the clear anterior view of the head.

We removed the plugs for 9-10 semi-randomly located brood cells (each group of 9-10 cells in a different colony = 1 biological replicate) and dispensed odorants (1 µl of neat solutions) onto the brood through the back of the comb and re-plugged each cell. The number of brood in these patches is smaller than for the front-way odorant assays because the size of the Jenter™ cage limits the total brood area. We traced a map of the odorant-treated cells and placed the combs in colonies for 20 h to allow workers to detect the odorant signals through the cap and respond. We performed five biological replicates (*i.e.* repeated the test in five colonies) for each odorant and developmental stage (pre-pupation and post-pupation). Since the 5th instar larvae and pre-pupae were too hard to confidently distinguish (as described above), the 'pre-pupation' group contains both stages. After incubation, we removed the comb and counted the number of brood cells from each odorant treatment that were removed and/or partially cannibalized. Removal data was analyzed as described above except we used a two-factor ANOVA (levels: odorant, age). Due to spontaneous re-queening events and subsequent worker turn-over, the hygienic scores are not known for all of the colonies in this experiment.

4.9.5 **Broodless back-way odorant assays**

These assays were performed exactly the same as the back-way assays containing brood, except the brood was first removed through the Jenter™ cell plug using clean forceps. Small strips of No. 1 Whatman filter paper were then inserted where the brood used to be, and 1 µl of odorant treatments (either neat or 1% v/v in hexane) were dispensed onto each filter paper. To try to increase the resolution of the assay, we used slightly larger patches of brood (13-15 cells, instead of 9-10). The Jenter™ cages were then incubated for 20 h and scored as before.

To produce live brood extracts, 15 prepupae were either crushed in 500 µl hexane with a glass rod, or washed in the same volume of hexane. For the crushed extract, the phases were allowed to separate for 10 min followed by extraction of the upper phase with a syringe. The hexane wash extract was produced by gently agitating the prepupae in the hexane for 5 min, followed by solvent extraction with a syringe. The solutions were stored at -20^{o}C until further use.

4.9.6 **Gas chromatography mass spectrometry sample collection and data analysis**

We performed GC-MS on extracts from larvae, prepupae and pupae to detect differences in their natural odorant profiles. Here, the three stages are distinguishable because by removing the brood from the cell, we can clearly differentiate the features of a prepupae compared to a 5th instar larva (the elongated body and 'crook-neck' appearance). We collected capped 5th instar larvae, prepupae and pupae from five different colonies and performed solid-phase micro-extraction (SPME) GC-MS as well as cuticle hexane wash GC-MS as previously described (Section 3.8.4). Data was analyzed as previously described except peak areas were log10 transformed and compound profiles were compared between developmental stages using a one-

way ANOVA followed by a Benjamini-Hochberg correction (5% FDR) performed in Perseus (v1.5.5.3).

4.9.7 **Electroantennography recordings**

EAG data acquisition was performed essentially as previously described (Section 3.8.5), but with minor changes to the odorant presentation method to make the data more statistically sound, and to include an odorant heating step. We heated the odorant presentation cartridges to 37°C using a flexible chromatography column heater, at which time we dispensed onto the filter paper 5 µl of distilled water (blank), β-ocimene, oleic acid, or a 1:1 v/v blend of β-ocimene and oleic acid (mix). After allowing 30 s of initial evaporation and slight cooling for the cartridge to reach approximately 33ºC, we aimed away from the antenna and passed a 1 s burst of room-temperature air through the pipette before stimulating the antennae with the odorants. We then exposed the antennae to a set of 3 consecutive 1 s bursts for each odorant in a randomly-determined order. Between 0.5 and 1 min was allowed between each presentation to allow antennal electrical activity to return to baseline. Blank stimuli (also 3 consecutive 1 s bursts each) were performed at two randomly determined times during acquisition. For each antenna, we subtracted the average blank intensity from the odorant EAG intensities, then compared odorant groups with a two-way ANOVA (levels: odorant, side).

4.9.8 **Odorant toxicity assay**

To test the toxicity of the odorants, we retrieved 60 purple-eyed, white body pupae and applied 1 µl of neat odorant (phenethyl acetate was not included) to the dorsal abdominal area (n = 15 each). We placed the pupae in tissue-lined petri dishes and incubated them at 33ºC for 2.5 d. We

then scored the pupae for whether their development was halted (*i.e.* their cuticle did not begin to brown or harden and their eye pigment did not change colour) and whether a prophenoloxidase response had initiated (*i.e.* the dorsal abdominal region became black). All pupae with halted development also had a prophenoloxidase response.

The following year, we repeated this toxicity assay with white-eyed, white bodied pupae to better match those that were used in the back-way odorant assays. This time, we monitored the pupae for prophenoloxidase responses after 1 d, and we included 1% odorant treatments in addition to 100% odorant treatments (15 pupae each).

Chapter 5: *In vitro* and *in vivo* systems for studying honey bee genes linked to hygienic behaviour[5]

5.1 Introduction and rationale

Hygienic honey bees have superior olfactory sensitivity compared to non-hygienic honey bees[91-94], which likely depends in part on differences in antennal gene expression[39,106,112,115,202]. In a search for antennal biomarkers for hygienic behaviour, we previously identified two odorant binding proteins – OBP16 and OBP18 – that significantly correlated with colony hygienic score[106]. Antennae are honey bees' main olfactory appendages, and OBPs aid odorant signal detection by binding and transporting hydrophobic odorant molecules from the antennal pores to the olfactory nerves[120]. As a biomarker, we know that the proteins' expression is tightly (positively) correlated with hygienic behaviour[106]; however, despite some tantalizing inferences, a mechanistic link between OBPs and hygienic behaviour has not yet been made.

The best way to experimentally determine if these OBPs are necessary for sensitizing bees to hygienic behaviour-inducing odorants is to alter their gene expression and observe the effect this has on their olfactory sensitivity. Insects are particularly amenable to gene knockdown via RNAi, and numerous publications have utilized RNAi to experimentally manipulate gene expression in honey bees. These experiments mainly targeted the fat body, embryos, parts of the brain, or viruses[58,165,240-248]; however, none have achieved knockdown in antennae flagella, which

[5] A portion of this chapter has been published: McAfee, A. *et al.* A death pheromone, oleic acid, triggers hygienic behaviour in honey bees (*Apis mellifera* L.). *Sci. Rep.* 8:5719. Copyright (2018) Authors. Permission not required for reprinting. This work is licensed under a Creative Commons Attribution 4.0 International License.

is our target. While long-lasting, systemic RNAi has been reported in other insects[249], to the best of our knowledge, it has not been reported in social insects; therefore, an RNAi response in honey bees may be transient, probably lasting in the order of days.

RNAi can be used to decrease gene expression, but in order to determine if OBPs are sufficient to sensitize bees to hygienic behaviour-inducing odorants, expression must be increased. Unfortunately, negative regulators of OBP16 and 18 expression (which could be targeted by RNAi in order to increase OBP expression) are not known, so another technique must be used. In 2014, Schulte *et al.* published the first ever method for producing transgenic honey bees using a piggyBac-derived cassette, which is an ideal technique for studying *in vivo* gene expression[250] and can be applied to this system to overexpress the OBPs. The same group has also characterized several honey bee promoter sequences, including one actin promoter and one neuron-specific promoter[251]. Since then, there have been no further publications utilizing these tools. Thus, using this technology, which is in its infancy, has a high risk of failure; however, if successful it could have great rewards. At a minimum, our work towards creating transgenic honey bees will help better establish this technique in the field and improve its utility for other researchers. Because of this risk, we also looked for ways to mechanistically link OBPs to hygienic behaviour without altering gene expression *in vivo.*

Odorant binding proteins have been the target of substantial *in vitro* investigations, including binding assays to test for their interactions with different ligands[106,122]. *In vitro* binding assays between OBP16, OBP18, and the candidate hygienic behaviour-inducing ligands identified in

Chapter 3 is therefore one more way that we can begin to investigate if they have the potential to be mechanistically linked, based on physical interactions.

5.2 Research goals

Our pinnacle goal is to experimentally manipulate expression of OBP16 and 18 in order to empirically determine the phenotypes dictated by these two proteins, particularly with respect to hygienic behaviour. We aim to achieve this using complimentary approaches: knocking down expression via RNAi and overexpressing the proteins via transgenics. Since RNAi has not been previously used to target gene expression in the antenna flagellum, we will attempt it via four different methods: 1) feeding dsRNA to larvae reared in the hive 2) feeding dsRNA to larvae reared *in vitro*, 3) feeding dsRNA to adults, and 4) directly injecting dsRNA into the flagella of pupae. To achieve overexpression, we aim to insert either one or both of the OBPs under the control of the honey bee actin promoter directly into the genome using the methods of Schulte *et al.*[250] We hypothesize that honey bees' sensitivity to disease odors should decrease with gene knockdown and increase with overexpression. If successful, this will be a major landmark in experimentally deciphering part of the molecular mechanism of hygienic behaviour.

Finally, we also perform *in vitro* ligand binding assays between four odorants (hexane, phenethyl acetate, oleic acid, and β-ocimene) and OBP16 and OBP18 to determine the relative strengths of these interacting partners. Taken together, the results of these experiments will let us gain insights and potentially propose a mechanistic model of how a co-opted, volatile brood pheromone (β-ocimene) could work together with an evolutionarily conserved death cue (oleic

acid) via interactions with hygienic behaviour-associated odorant binding proteins (OBP16 and OBP18) to induce hygienic behaviour.

5.3 *In vitro* ligand binding assays

To test if β-ocimene and oleic acid are strong binding partners for hygienic behaviour-associated OBPs, we performed *in vitro* binding assays with OBP16 and OBP18 using the NPN (N-phenyl-1-naphtylamine) competitive displacement approach. This assay measures the fluorescence emitted when NPN binds the OBP, and the fluorescence decrease that occurs when an odorant molecule of interest out-competes NPN for the binding site. Like our front-way behavioural assays, we used hexane as the odorant negative control and we included phenethyl acetate despite the surprising outcomes of behavioural tests (depicted in Figure 4.2). We found that of the four tested odorants, hexane and phenethyl acetate did not strongly displace NPN from its binding site to OBP16 nor OBP18 (Figure 5.1). These compounds were also poor inducers of hygienic behaviour, as determined by our previous behavioural assays (Figure 4.2 B). β-ocimene, however, displaced NPN from OBP16, but less so from OBP18. Oleic acid displaced NPN from both OBPs. This data indicates that β-ocimene could be a ligand for OBP16, and oleic acid could be a ligand for both OBPs.

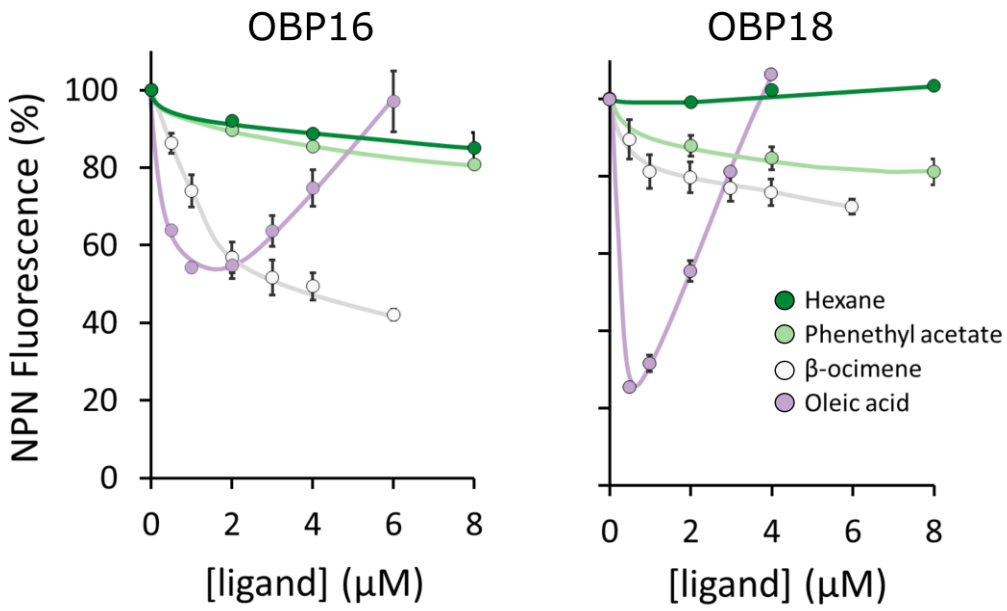


Figure 5.1. *In vitro* NPN competition assay for OBP16 and OBP18

We used an NPN (N-Phenyl-1-naphthylamine) competitive binding assay to measure affinities of β-ocimene, oleic acid, phenethyl acetate, and hexane (negative control). Assays were performed in technical duplicate with 2 µM protein and 2 µM NPN in all cases. Lower NPN fluorescence intensity indicates stronger ligand binding. The high NPN fluorescence intensity for the high oleic acid concentrations is due to the formation of micelles at higher concentrations of the ligands. A 1% solution of β-ocimene, oleic acid, phenethyl acetate, and hexane corresponds to approximately 60 mM, 32 mM, and 63 mM, and 76 mM, respectively. Error bars are standard error of the mean.

5.4 RNA interference against odorant binding proteins

In insects, RNAi is typically achieved by either feeding, soaking, or injecting the organism with long dsRNA (>300 bp) that shares its sequence with the target gene[249,252]. Getting the dsRNA to the desired organ or cells is the first challenge in any RNAi experiment. Here, we attempt to achieve knockdown by three different delivery methods: feeding and soaking, feeding alone, and microinjecting dsRNA directly into the antennae.

5.4.1 **Delivery by feeding to hive-reared larvae**

Since dsRNA uptake by insects has been reported via both feeding and soaking[249], our first approach was to mix dsRNA with the food of young larvae. Since the larvae are in constant contact with the liquid food, they will both feed on the dsRNA and bathe in it, so they may take up the dsRNA through both routes of entry. Larvae can be grafted and reared to adulthood in the laboratory[253-255]; however, this is a tedious and time-consuming process. Moreover, it can result in physiological changes in the adult bees[254]. Therefore, we chose to add dsRNA directly to the pool of each developing larva's food (1 dose of 1 µg per larva) within a naturally laid brood comb. We treated 2nd instar larvae with either OBP16, OBP18, or GFP dsRNA, and returned the frame to the hive until the cells were capped. Once emerged, we harvested the adult bees and quantified OBP16 and OBP18 levels in their antennae using a multiple reaction monitoring (MRM) approach (Figure 5.2). While the results were promising, they were highly variable.

The GFP treatments were especially variable, such that there were no significant differences between GFP and any other treatments (two-way ANOVA). The OBP16 dsRNA tended to decrease OBP16 abundance, but surprisingly, the greatest decrease was actually observed for OBP18. This is surprising, but not outside feasibility – OBP16 and OBP18 share a lot of sequence similarity and while we designed the dsRNA to limit cross-reactivity, this data suggests that could still be occurring. The OBP18 dsRNA showed the expected expression patterns: no decrease in OBP16 expression and a modest decrease (~50%, although this is not a significant difference) in OBP18 expression. Importantly, OBP21 expression (which was not targeted) remained approximately the same across treatments. Overall, the percent knock-down was not

sufficiently strong for either of the treatments and the results were too variable. We therefore sought a method of achieving knock-down that yields stronger and more reliable results.

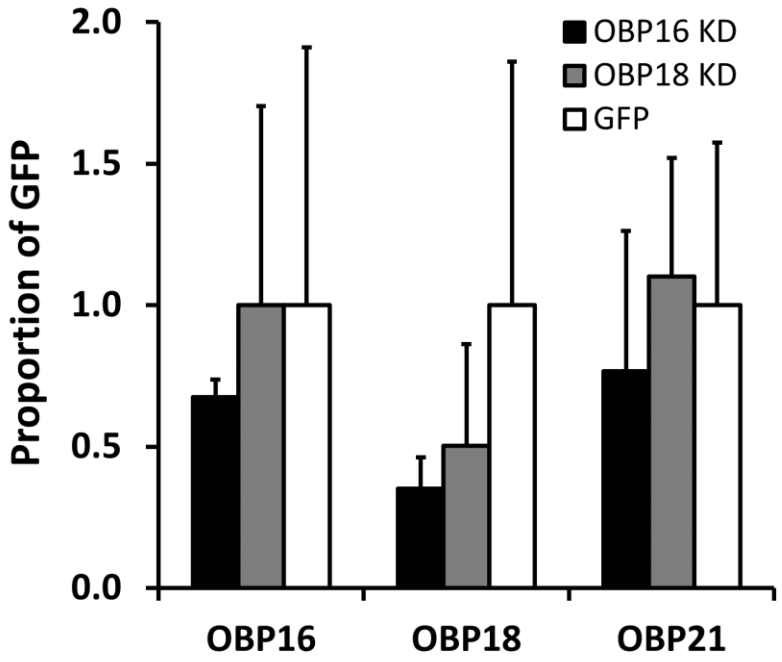


Figure 5.2. Results of RNAi experiments with dsRNA delivery via in-hive feeding.

Larvae were supplied with a single dose (1 µg) of either OBP16, OBP18, or GFP dsRNA delivered by pipetting 1 µl of dsRNA solution into the food droplet of 2nd instar larvae. Three patches of ~20 brood cells were treated per type of dsRNA within a single colony. Patch locations were drawn on a plastic transparency and adult bees were harvested upon emergence. Antennal protein was quantified by multiple reaction monitoring mass spectrometry. Data was normalized to the average abundance in the GFP dsRNA treatments. There were no significant differences in protein expression (two-way ANOVA; levels: treatment, protein). Error bars represent standard deviation.

5.4.2 **Delivery by feeding *in vitro*-reared larvae**

Our second approach was to feed dsRNA to larvae reared *in vitro*. While more time consuming, this method offers a more controlled rearing environment and is more amenable to sustained dsRNA treatment over the course of several days. We reared larvae in excess brood food

containing dsRNA (50 µg per 1 mL of food) for three days (days 5 through 8 post-laying, *i.e.* the 2nd to 5th instars) then allowed the larvae to continue development without dsRNA treatment. In a preliminary experiment to obtain survival curves (and therefore gauge how many larvae we needed to begin with for each replicate) we fed groups of 30 larvae either no dsRNA, GFP dsRNA, or a 1:1 w/w mix the OBP16 and OBP18 dsRNA (Figure 5.3). Approximately 50% of larvae in all treatment groups reached pupation; therefore, beginning with 30 larvae per replicate yielded enough material for antennal proteomics samples.

We repeated the experiment above but applied OBP16 and OBP18 dsRNA individually, rather than as a mix. After quantifying the OBPs using the MRM method as in Section 5.4.1, we found that the expression levels were highly variable again (Figure 5.4). In the antennae, there were no statistically significant differences; moreover, the biggest difference in expression did not follow the same pattern as before (*i.e.* OBP16 dsRNA affecting OBP18 expression). We also analyzed protein expression in the legs, which also express these OBPs and should be equally affected by these dsRNA treatments. There, we found that again OBP16 dsRNA appears to unexpectedly affect OBP18 expression (two-way ANOVA followed by Tukey HSD test; $p < 0.05$). However, we also found that OBP16 dsRNA affects expression of OBP21, which we were not targeting in this experiment, and which was not affected in the hive-reared larvae. Overall, these results are not consistent nor specific enough to proceed with behavioural or physiological assays.

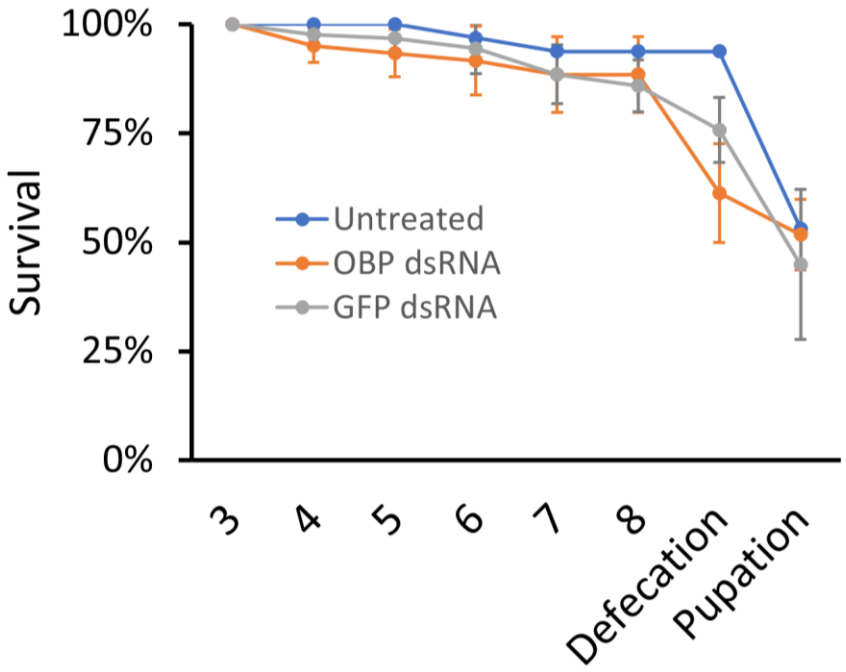


Figure 5.3. Example survival curves for in vitro rearing.

Numbers on the x-axis refer to days post-laying. Newly hatched larvae were grafted into culture plates containing excess larval food. On day 5 post-laying (*i.e.* during the larvae's 2nd instar), the food was mixed with either no dsRNA, GFP dsRNA or a 1:1 mix of OBP16 and OBP18 dsRNA (50 µg/mL, n = 30 larvae per replicate, N = 4 replicates for OBP and GFP treatments, N = 1 for untreated larvae). Larvae were maintained on dsRNA supplemented food (replaced daily) for 3 days, then completed development without dsRNA treatment.

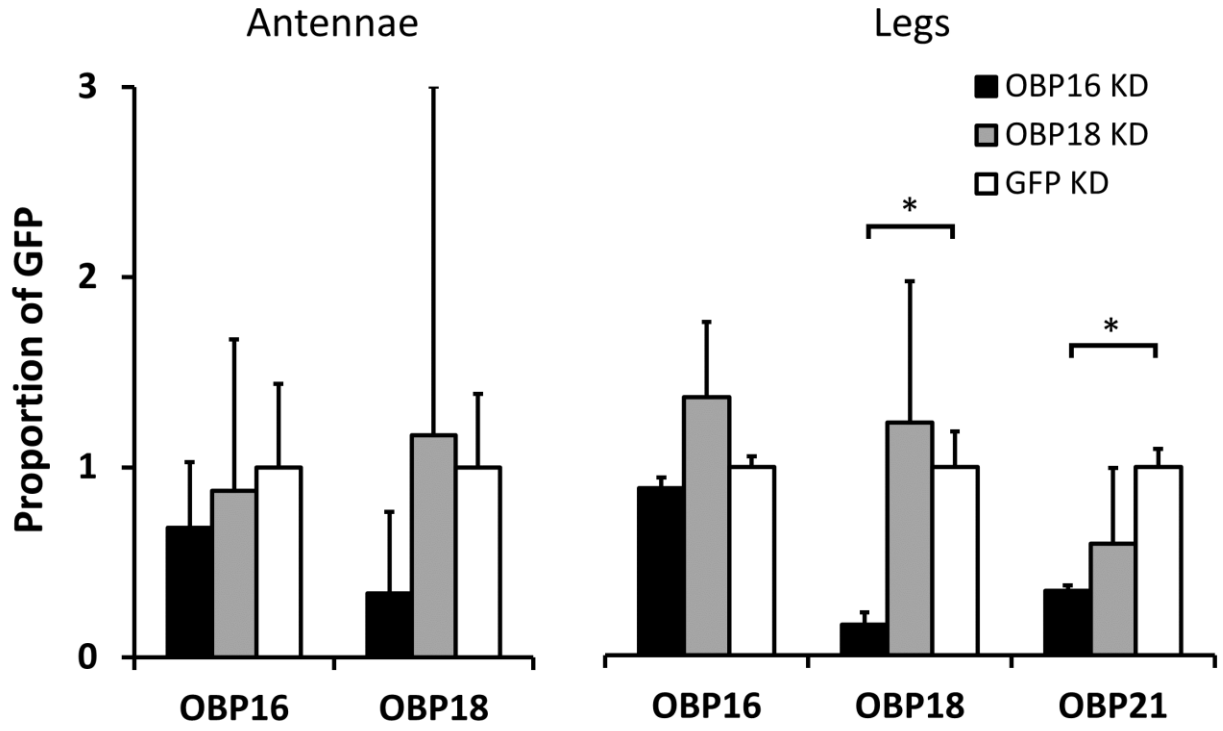


Figure 5.4. Results of RNAi experiments with dsRNA delivery via *in vitro* larval feeding.

Larvae were fed either OBP16, OBP18, or GFP dsRNA (50 µg/mL, n = 30 larvae per replicate, N = 4 replicates per treatment). 2nd instar larvae were maintained on dsRNA-supplemented food for 3 days, then completed development without dsRNA treatment. Adults were harvested once they were able to walk. We were not able to detect OBP21 in the antennae in this experiment. Protein was quantified by multiple reaction monitoring mass spectrometry and statistics were performed using a two-way ANOVA (* indicates $p < 0.05$). Error bars are standard deviation.

5.4.3 **Delivery by adult feeding**

We reasoned that the lack of consistency and specificity of dsRNA response could be due to batch variation of the *in vitro* dsRNA synthesis reaction or decay of the RNAi response over time (there is approximately 13 days between the last dsRNA treatment and adult emergence), or both. To improve quality of the dsRNA, we re-designed the dsRNA molecules to be longer (via repeats of the short unique regions within each OBP gene) and purchased the custom-synthesized dsRNA from a biotechnology company (AgroRNA, Seoul, South Korea). Previously, the OBP16

and OBP18 dsRNA molecules were 95 and 110 bp, respectively. The newly designed dsRNAs are 454 and 448 bp, respectively (GFP remained the same, at 252 bp). Before using the new source of dsRNA in experiments, we confirmed its activity by transfecting *D. melanogaster* S2 cells with a GFP expression plasmid, then using the GFP dsRNA to knock down expression (Figure 5.5). Once activity was confirmed, we then used the dsRNA to feed adult workers (50 µg per 1 mL of 50% sugar syrup) with either GFP dsRNA or a 1:1 w/w blend of the OBP dsRNAs, since we observed cross-reactivity in our previous experiments. Unlike larvae, adults' antennae can be harvested sooner after dsRNA treatment, avoiding signal decay over time. After feeding workers the dsRNA-containing sucrose solution for 5 d, we harvested the bees and again quantified OBP expression in the antennae via MRM assays (Figure 5.6). Unfortunately, we found no significant differences in expression of any OBPs.

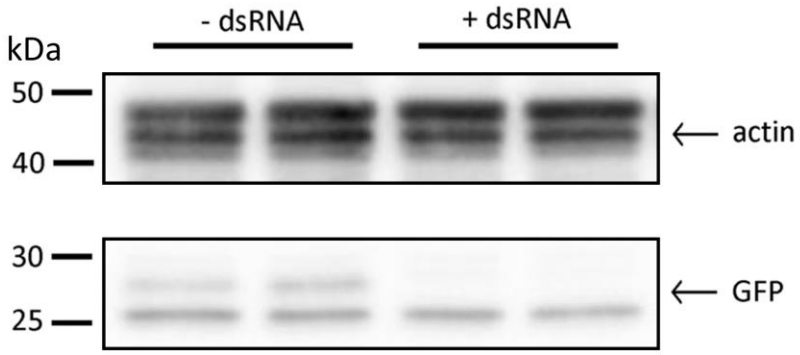


Figure 5.5. Western blot confirming GFP dsRNA activity.

***Drosophila* S2 cells were transfected (XtremeGene) with a GFP expression plasmid followed by a GFP dsRNA treatment via media soaking or a mock treatment (n = 2). Cells were incubated overnight before harvesting. We loaded 20 µg of protein in each gel lane and co-probed for GFP (27 kDa) and beta actin (42 kDa; loading control) before secondary incubation and HRP imaging. Primary antibody sources were DSHB (GFP-1D2) and Abcam (ab8224).**

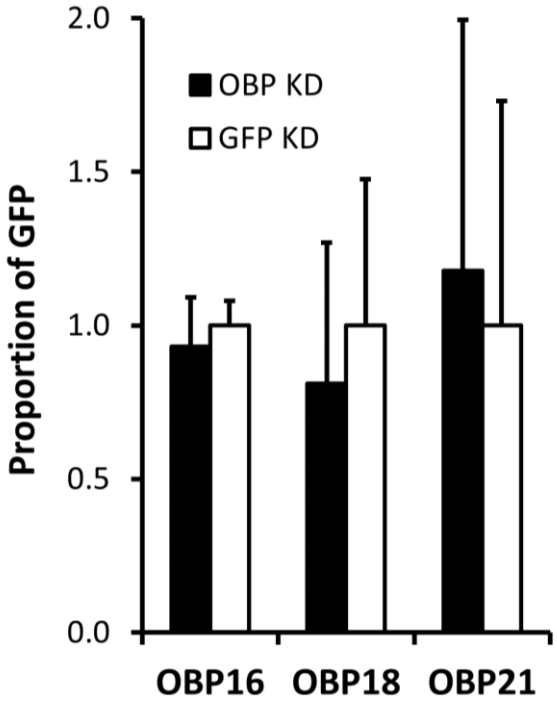


Figure 5.6. Results of RNAi experiments with dsRNA delivery via adult feeding.

Newly emerged adult worker bees were kept in plastic cages (n = ~20 bees per cage; N = 5 cages per treatment) in a humid incubator. Bees were fed an excess of 50% sugar syrup containing dsRNA at 50 µg/mL. After five days of treatment, antennal OBP expression was analyzed by multiple reaction monitoring mass spectrometry. "OBP KD" refers to a treatment composed of a 1:1 w/w mix of OBP16 and OBP18 dsRNA. Error bars represent standard deviation.

5.4.4 **Delivery via antennal microinjection**

To alleviate any doubt over whether the dsRNA was arriving at the target organ (the antennae flagella, we also delivered dsRNA via direct antennal microinjection. We injected 0.1 µg of either OBP18 dsRNA, GFP dsRNA, or no RNA into each flagellum of pink-eyed worker pupae (14-15 pupae per replicate, 3 replicates per treatment). At this stage, the cuticle has not yet hardened and is still translucent, allowing for facile injections with glass needles. By this method, we achieved very high survival (Table 5.1). We then analyzed protein expression by both MRM and shotgun LFQ proteomics in order to determine both the OBP expression levels as well as any off-target effects induced by injection (Figure 5.7). We tested two types of dsRNA:

one that underwent a new, proprietary purification method (also from AgroRNA; labeled "New" in Figure 5.7 A) and one that did not (*i.e.* the same dsRNA that was used in the adult feeding trial). In this experiment, we only tested GFP and OBP18 dsRNA, since we could only receive two samples of the newly purified product. Unfortunately, the results are again inconclusive. We see opposite patterns of expression for OBP18 and OBP21 between the two dsRNA batches, with a general trend for the "New" dsRNA to induce more knock-down, even in the GFP dsRNA negative control. Furthermore, upon shotgun proteomics analysis, it is clear that just performing an injection, whether with GFP or OBP18 dsRNA, causes a dramatic proteomic shift compared to the untreated controls (Figure 5.7 B). For these reasons, we do not think that direct antennal injection is a suitable method of dsRNA delivery.

Table 5.1. Pupae survival after antennal injections

Treatment	Surviving	Total
Untreated	15	15
Untreated	15	15
Untreated	15	15
OBP18	15	15
OBP18	14	15
OBP18	15	15
GFP	15	15
GFP	14	14
GFP	13	14

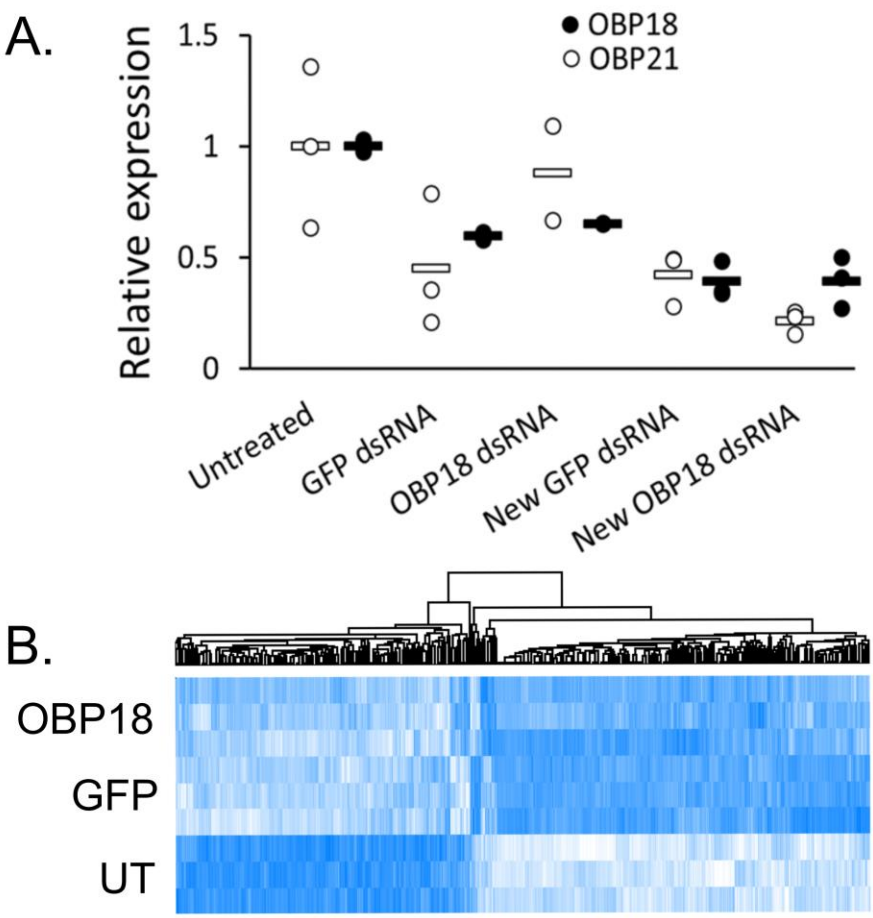


Figure 5.7. RNA interference against OBP18 in the antennae.

A. We microinjected dsRNA (GFP: 252 bp; OBP18: 454 bp) into the flagella of pink-eyed pupae and allowed them to reach adulthood (~6 d). Untreated bees received no dsRNA. GFP and OBP18 dsRNA bees received the GFP or OBP18 sequence (synthesized by AgroRNA but purified in-house), and New GFP and OBP18 dsRNA bees received the same GFP or OBP18 dsRNA sequence, but which underwent a proprietary purification method by AgroRNA. We then extracted the antennal proteins and quantified OBP18 and OBP21 peptides using multiple reaction monitoring mass spectrometry (Agilent 6460 QQQ). B. Heatmap of shot-gun proteomics analysis comparing microinjected antennae. We identified 2,852 proteins (25,640 unique peptides, 1% peptide and protein FDR) using a Bruker Impact II QTOF mass spectrometer coupled to an EASY-nLC 1000. This heatmap shows only the 918 proteins that were differentially expressed (label-free quantitation, one-way ANOVA, Benjamini-Hochberg corrected 5% FDR). Z-score scale: white = -2.5, blue = +2.5. Hierarchical clustering was performed in Perseus using average Euclidian distance (300 clusters, maximum 10 iterations). UT = uninjected. OBP18 = antennae were injected with OBP18 dsRNA. GFP = antennae were injected with GFP dsRNA.

5.5 Overexpression of odorant binding proteins

Since all four RNAi experiments failed to produce a knock-down response of sufficient quality to proceed with physiological or behavioural assays, we turned to developing transgenic methods to overexpress the OBPs. In these experiments, we utilize three different plasmids (Figure 5.8), all derived from the piggyBac transposon system. The first plasmid was kindly supplied by Dr. Christina Schulte and Dr. Martin Beye, and is identical to the plasmid used in their proof-of-principle report of transgenic honey bee rearing[250]. This plasmid contains the honey bee actin promoter driving GFP expression. The second plasmid is identical to the first, except the GFP gene is replaced with the OBP18 gene. Finally, the third plasmid uses the actin promoter to drive expression of a polygenic concatemer, with the genes for OBP16, OBP18, and GFP separated by a T2A viral self-cleaving peptide sequence. In addition to these elements, all three plasmids also contain the rubia fluorescence reporter gene driven by a neuron-specific promoter (6XP3).

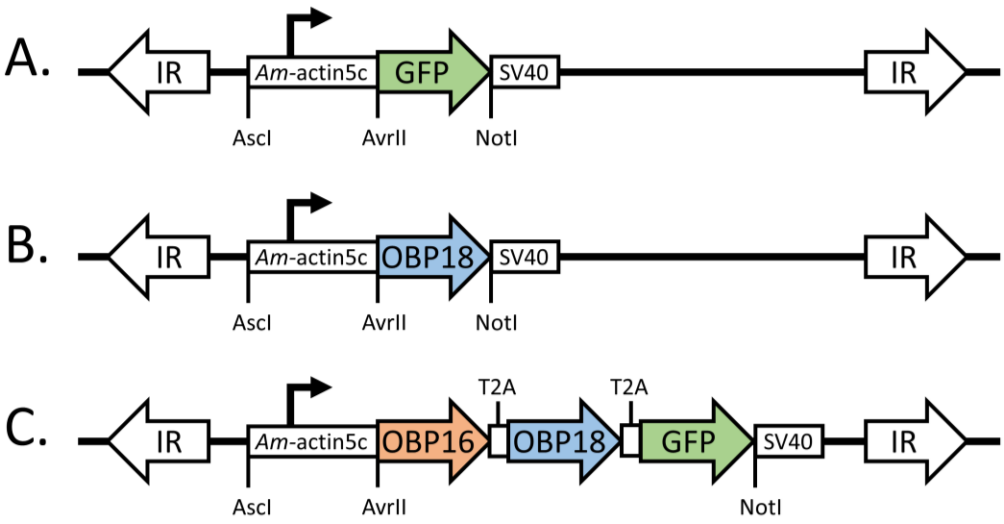


Figure 5.8. Schematic of transposon-containing plasmids.

All plasmids are derived from the piggyBac backbone. *Am*-actin5c is a honey bee actin promoter, SV40 is a transcription termination signal, T2A is self-cleaving peptide signal, and IR stands for "inverted repeat". All plasmids also contain the rubia red fluorescence reporter gene driven by a neuron-specific promoter (6XP3) which is functional in honey bees.

5.5.1 *In vitro* verification of the transgene cassettes

In order to confirm that the transgene cassettes can facilitate gene expression, we transfected *D. melanogaster* S2 cells with either plasmids A, C, or no plasmid (mock – transfection reagent only). Both plasmids A (pBac) and C (T2A) induced low levels of GFP expression, as determined by Western Blot (Figure 5.9 A). In both cases, GFP was detected at the correct molecular weight. This is important because since plasmid C is composed of a polygenic concatemer, with GFP as the terminal gene, this means that the T2A self-cleaving peptide sequence is at least partially functional, liberating the correctly-sized GFP. Otherwise, GFP would appear to have a molecular weight shifted to 15 or 30 kDa heavier, depending on how many T2A cleavages failed. However, since the expression was weak to begin with, less-abundant bands from missed T2A cleavages would probably not be visible, if present.

GFP expression from these plasmids was confirmed by fluorescence imaging of S2 cells after transfection (Figure 5.10). In this experiment, we included a positive control plasmid that carries GFP driven by a *D. melanogaster* actin promoter. GFP expression was most pronounced in the positive control, but was still present (albeit diminished) in the pBac and T2A transfections, and was not detectable in the negative control (mock – reagent alone).

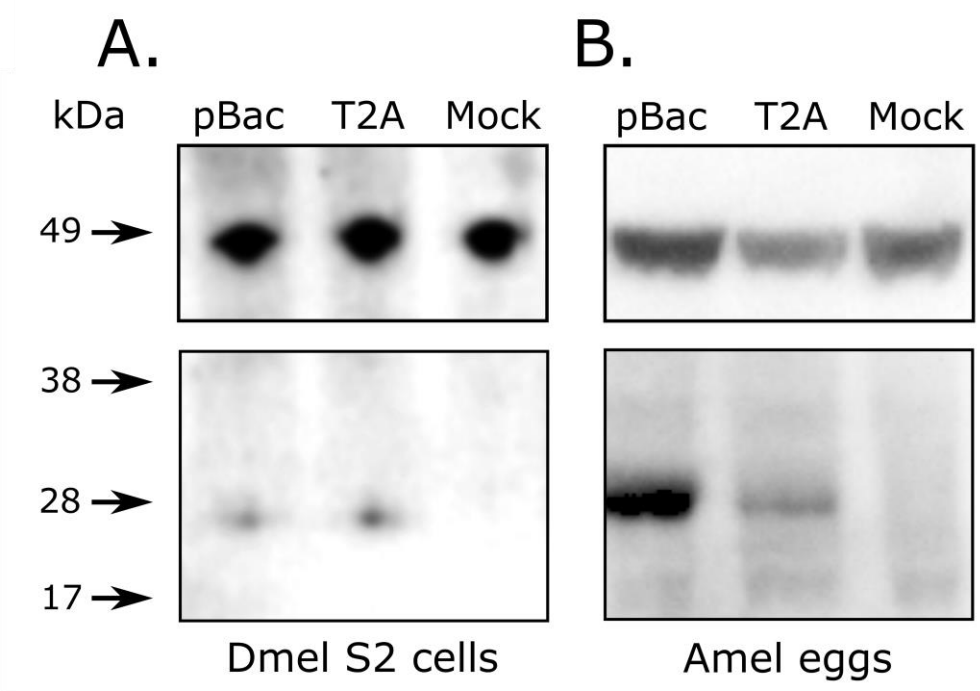


Figure 5.9. Western blot analysis to test plasmids for GFP expression.

A. *D. melanogaster* S2 cells transfected with GFP-containing expression plasmids. pBac and T2A refer to plasmids A and C, respectively (as shown in Figure 5.8), and mock refers to S2 cells treated with transfection reagent (XtremeGene) alone. Cells were incubated for 2 d before harvesting, then 50 µg of extracted protein was loaded in each lane and analyzed by the same procedure as described in Figure 5.5. B. Fresh honey bee eggs (0-2 h old) were collected and microinjected with the same plasmids as in A along with transposase for genomic integration. Mock refers to injections with plasmid B, which does not contain GFP. Eggs were incubated at 35ºC and after 2.5 d, 100 µg of extracted protein was loaded in each lane. GFP (27 kDa) and beta tubulin (50 kDa). Antibody sources were DSHB (GFP: 1D2; tubulin: E7-s).

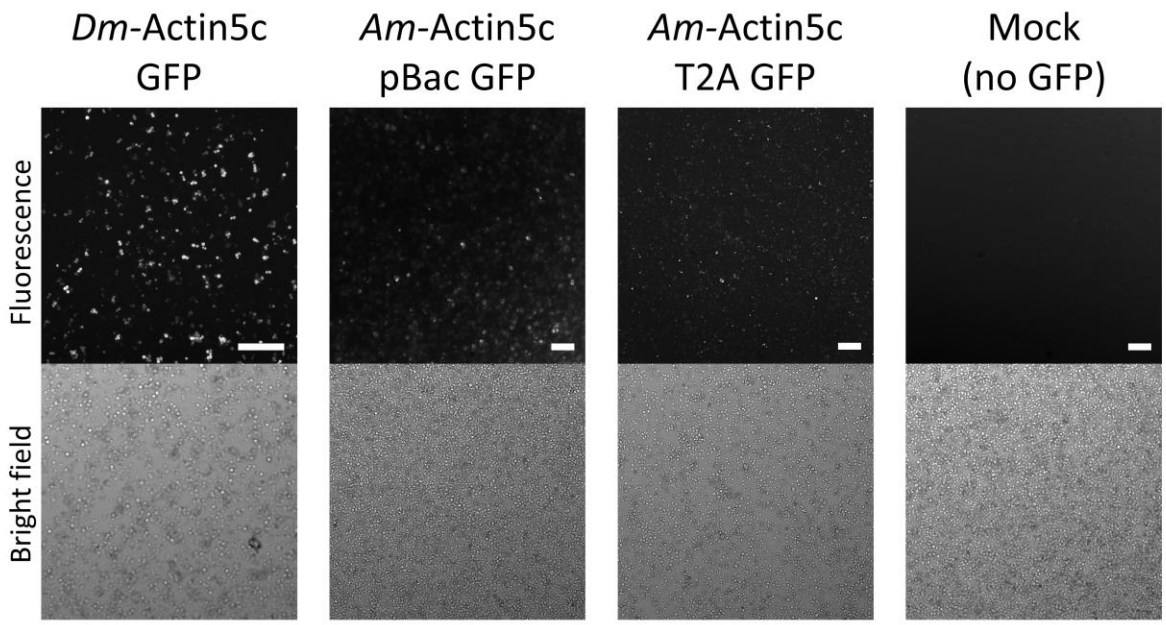


Figure 5.10. Analyzing S2 cell GFP expression by fluorescence microscopy.

***D. melanogaster* S2 cells were transfected (XtremeGene) with three different plasmids containing GFP: the positive control (far left) contained GFP under the control of a *D. melanogaster* actin promoter, while in the other plasmids GFP was driven by the honey bee actin promoter (*Am*-Actin5c). These plasmids correspond to A and C in Figure 5.8. Cells were incubated for 2 d before imaging. Scale bars represent 100 µm.**

5.5.2 *In vivo* expression analysis

Since both plasmids A and C utilize a honey bee promoter, it is possible that expression levels in S2 cells were low because we are using a heterologous cell line. Therefore, we also sought to confirm expression in honey bee eggs and larvae. We microinjected 0 to 2 h old honey bee eggs with either plasmid A, B, or C (in this case, plasmid B, which doesn't carry GFP, is the mock injection) along with transposase mRNA for cassette insertion into the genome. After 2.5 d we harvested the injected eggs and checked for GFP expression by Western blot (Figure 5.9 B). We found that the pBac-injected eggs (plasmid A) expressed GFP more strongly than the T2A-injected eggs. This may be because the T2A-containing cassette (which contains 4 genes) is

much longer than the pBac cassette (which contains 2 genes), and genomic integration events seem to negatively correlate with cassette size (Martin Beye, personal communication). No GFP was observed in protein extract from eggs injected with plasmid B (the non-GFP, mock plasmid).

We allowed a subset of the eggs to hatch into larvae, which we immediately imaged using fluorescence microscopy (Figure 5.11). Both the pBac and T2A-injected larvae emitted a strong fluorescence signal, while the mock-injected larvae (receiving plasmid B) emitted a weak signal, which is presumably autofluorescence. Honey bee larvae are notoriously prone to emitting strong autofluorescence signals, particularly upon tissue damage or death (Martin Beye, personal communication); therefore, we caution against overinterpreting GFP fluorescence images. However, combined with the prominent GFP band in the Western blot, we are confident that GFP is being expressed.

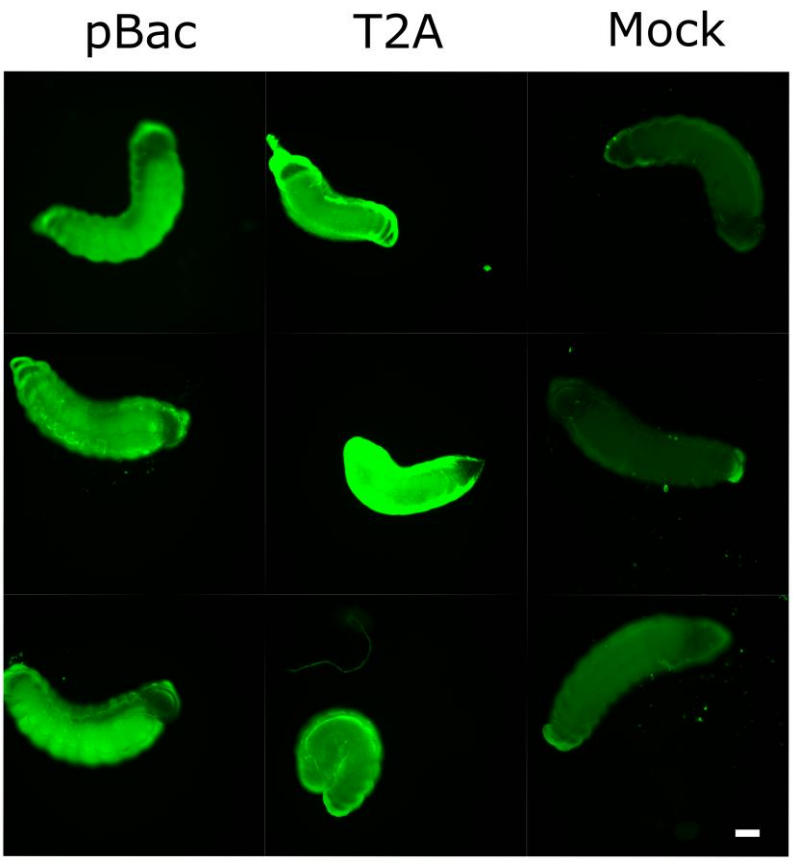


Figure 5.11. *In vivo* confirmation of GFP expression.

Fresh honey bee eggs (0 to 2 h old) were injected with either plasmid A, B, or C, where B is the non-GFP negative control (mock; described in Figure 5.8). Larvae were incubated at 35ºC for 3 d and imaged by florescence microscopy immediately after hatching. Exposure time was set to 0.1 s for all images. The scale bar represents 100 µm. Only a small fraction of the pBac- and T2A-injected larvae emitted a fluorescent signal. Images were obtained using a Cellomics high-content imager (filter: 485_BGS_BGS)

5.5.3 **Improving egg hatching rates**

One challenge we encountered in our attempts to produce transgenic honey bees is that very few of the injected eggs actually hatch. In 2017, our hatching rates varied from 1% to 5%, which is far lower than what Schulte *et al.*[250] achieved (45%). We injected young (0 to 2 h old) eggs to maximize the chances that an integration event would lead to an adult carrying the transgene in her germ line, but we wondered if the tradeoff between injecting older eggs, which contain more

complex multicellular embryos, and potentially achieving a higher hatching rate might be a more economical approach. We therefore injected eggs harvested after caging the queen for 3 h, 6 h, and 16 h with the pBac plasmid (plasmid A). We found that eggs harvested after 3 h and 6 h hatched at similar rates (about 5%), whereas eggs harvested after 16 h hatched at much higher rates (about 50%) (Figure 5.12 A).

The concern over injecting older eggs is that since the embryo is composed of many cells at that age, it is less likely that an integration event will occur in the cells that eventually go on to produce the germ line. To roughly gauge integration heterogeneity, we performed fluorescence microscopy on larvae that were injected as 16 h old eggs (Figure 5.12 B). We imaged the red fluorescence channel corresponding to rubia, rather than GFP, because honey bee larvae emit less autofluorescence at these wavelengths (Martin Beye, personal communication). We found that some of the injected larvae indeed had patches of cells emitting a red fluorescent signal, which was not observed in the control larvae. Here, control larvae were not injected with any plasmid, since all the piggyBac plasmids contain the rubia reporter gene.

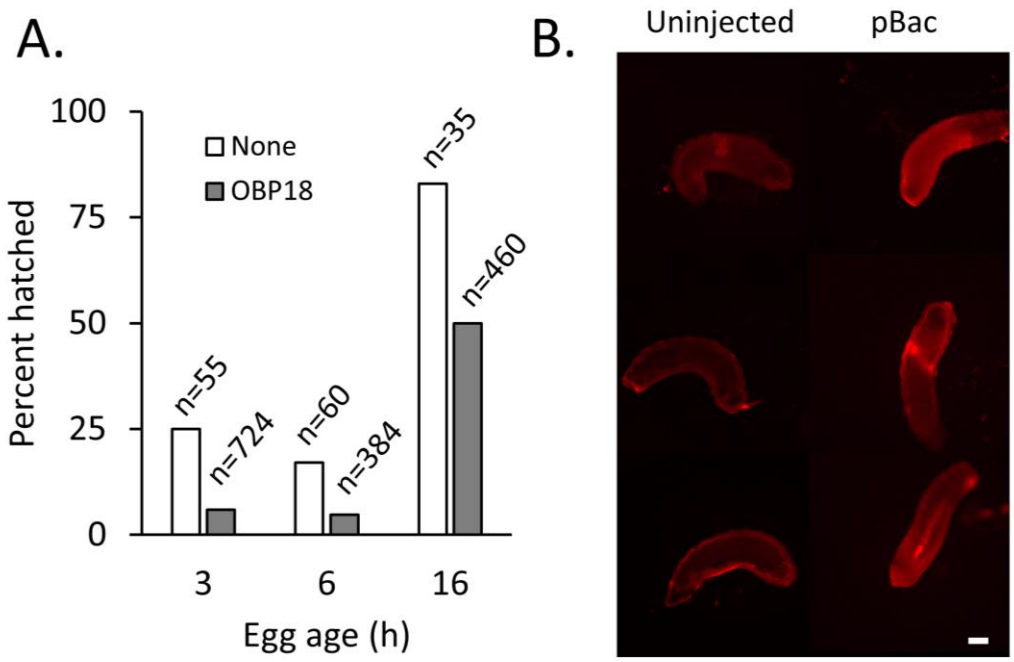


Figure 5.12. Investigating the effect of injecting older eggs

A. Honey bee eggs were obtained from queens that were caged in a Jenter egg collection chamber for 3 h, 6 h, and 16 h. Uninjected eggs represent handling controls. Injected eggs received the OBP18 plasmid (plasmid B). B. In a separate experiment, we injected 16 h eggs with the pBac plasmid (plasmid A) and performed fluorescence microscopy on the larvae immediately after hatching. The scale bar represents 100 µm. Images were obtained on a Cellomics high-content imager (filter: 549_BGS_BGS). Exposure time was set to 0.1 s for all images.

5.6 Conclusion and future directions

The experiments described here clearly do not meet our objectives of being able to perform behavioural or physiological tests on honey bees with altered OBP16 or 18 expression. However, we were able to gain some new insights into these OBPs' mechanistic role through *in vitro* binding assays. The RNAi experiments offer a cautionary tale for future work investigating antennal gene expression, and the transgenic plasmid expression confirmation represents a significant step towards creating transgenic honey bees expressing these genes.

5.6.1 *In vitro* binding assays

With a large body of research showing that olfaction is important for hygienic behaviour, combined with two odorant binding proteins (OBP16 and OBP18) emerging as protein biomarkers, a tempting hypothesis is that the OBPs are aiding the detection of odorants associated with disease or death. The NPN competitive binding assay does not directly measure odorant affinity, since we cannot be sure that NPN binds to the OBP in the protein's native binding pocket or somewhere else on the molecule. Furthermore, what we measure is essentially the odorant's ability to displace NPN, rather than the odorant's direct affinity to the OBP's binding pocket. Nevertheless, some research groups use this information to calculate a provisional binding constant[106,256].

Here, we are using the NPN assay to try to approximate the relative interaction strengths between the four candidate ligands (hexane, phenethyl acetate, β-ocimene, and oleic acid) and OBP16 and OBP18. β-ocimene is a good candidate ligand for OBP16. Oleic acid is a good candidate ligand for both OBP16 and OBP18, as it was the best at displacing NPN of all the ligands we tested. Since β-ocimene and the odorant blend induced significantly higher antennal nerve depolarizations than oleic acid (Figure 4.8), this suggests that either the worker bees must be very close to the emanating cell (or possibly even contacting the source) to sense it, or the odorant treatment induces the brood to emit a different, more volatile signal. Nevertheless, the observation that hygienic behaviour-inducing odorants physically interact with these OBPs more strongly than odorants that are poor hygienic behaviour inducers supports the OBPs central role in enabling the behaviour.

While oleic acid produces high fluorescence intensities (which normally indicates weak binding) at higher ligand concentrations (*i.e.* > 1 µM), this is a well-known phenomenon for amphipathic ligands[257,258]. The very low fluorescence intensity < 1 µM indicates that it is indeed a strong NPN displacer for OBP18, which agrees with previous binding assays[106]. Conversely, the two odorants which induced low rates of hygienic behaviour in our assays also were poor ligands for these OBPs. Therefore, the results of this *in vitro* binding assay can explain the behavioural observations surprisingly well.

Despite this evidence, it is difficult to know how well the OBP and ligand concentrations reflect reality. For example, the absolute concentration of OBPs in the hemolymph of honey bee antennae is currently unknown, as is the effective ligand concentration at the antennal pore (the interface between the hemolymph and the surrounding air). While a 1% solution of β-ocimene corresponds to approximately a 60 mM solution, which is much higher than the concentrations in the ligand binding assays (<10 µM), with volatility mechanics and spatial diffusion, the airborne concentration is likely much lower.

5.6.2 **RNAi**

RNAi is a widely-used technique, but it can produce a notoriously variable responses depending on the gene, tissue, and siRNA sequences within the longer dsRNA. Performing RNAi against odorant binding proteins has the added challenge of finding a sequence that is both long enough to induce efficient RNAi, and does not contain cross-reactive siRNAs. This is because honey bee OBPs have remarkably similar sequences to one another; however, they are sufficiently functionally divergent for only two of them (16 and 18) to correlate with hygienic behaviour.

The antennae, too, are a challenging organ to target because they are located distant from the point of dsRNA uptake (typically the midgut, but possibly also the larval cuticle via the soaking method). We expected that the dsRNA may be taken up through the midgut and transported to the antennae via the hemolymph; unfortunately, from our results, we are unable to tell if the failure to decrease gene expression was a result of the dsRNA not arriving at its target location, or the dsRNA itself being ineffective. We hoped to clarify this with the direct antenna injections; however, the sheer magnitude of the proteomic shift after injection (908 proteins were differentially expressed) leads us to conclude that any change in OBP expression (or lack thereof) comes with too many other changes in expression for us to reliably interpret it as being a direct result from our dsRNA treatment. Throughout our experiments, we were unable to find a combination of delivery methods, time frame, and dsRNA sequence that would allow for efficient, reliable gene knockdown of our targets.

5.6.3 **Transgenics**

After the failure to decrease OBP gene expression, we instead turned to transgenics to focus on increasing gene expression. This is an arduous technique that is not yet widely used in the field; however, it also offers a great opportunity. Once produced, a transgenic queen can offer a sustained source of workers that are overexpressing the OBPs, allowing their effect on both individual and colony-level phenotypes to be determined. Here, we have not yet made a transgenic queen, but we have confirmed that our expression plasmids function as expected when they are injected into honey bee eggs via Western blot and fluorescence microscopy of larvae. This is significant because while the T2A cleavage sequence has been shown to work in other insects such as *Bombyx mori*[259] and *D. melanogaster*[260] as well as zebrafish and mouse cells[261], it

has not yet been demonstrated in honey bees. Here, we were able to detect GFP at the expected molecular weight (27 kDa) even when located downstream of a T2A sequence, indicating that it has been liberated from the upstream genes. Expression was above the limit of detection but still weak; therefore, further experiments confirming these results will be required. For instance, since expression of the T2A-containing cassette was not strong enough to observe the GFP which was not liberated from either of the upstream genes (*i.e.* a ~59 kDa GFP signal), or GFP which may be linked to OBP18, but liberated from OBP16 (*i.e.* a ~43 kDa GFP signal), we are unable to determine the cleavage efficiency of this system.

The next step is to rear plasmid- and transposase-injected larvae into queens, using standard procedures in beekeeping, and screen those queens for ovary incorporation (via PCR screening her drone sons). Semen can then be collected from transgene-positive drones and used to instrumentally inseminate other queens, creating a fully functional colony for further experimentation. We expect that the incorporation efficiency of the T2A-containing plasmid may be low, since it is ~850 bp longer than the plasmid containing GFP alone, and anecdotally, incorporation efficiency appears to be negatively correlated with insert length (Martin Beye, personal communication). Furthermore, the T2A sequence leaves an 18 amino acid C-terminal sequence tagged to the upstream protein (*i.e.* OBP16 and OBP18) as well as an N-terminal proline to the downstream protein (only affecting OBP18). The effect that this may have on OBP functionality is unknown. It is possible that these sequence remnants could change the shape of the binding pocket or, in the case of the C-terminal tag, occupy the binding pocket itself. For these reasons, we are not proceeding with the T2A-containing cassette, but a conventional cassette containing the OBP18 sequence alone driven by the *Am*-Actin5c promoter.

OBP18 is expressed in the goblet cells of olfactory sensilla on the antennae, but it is also expressed in several other tissues (cuticle, legs, brain, tergite, and sternite)[120]. The *Am*-Actin5c promoter, however, will drive expression throughout the body, and the physiological effects of this widespread expression are unknown. While it is an imperfect overexpression strategy, given the promoters that have been characterized for honey bees, it is likely the best option. Schulte *et al.*[251] have done by far the most work on sequencing and characterizing honey bee promoters. They have established not only the *Am*-Actin5c promoter, but also elp2l (neuron-specific), as well as *Am*-hsp70 and *Am*-hsp83 promoters. They have also shown that the 6xP3 promoter can induce expression in the honey bee compound eye. Their work shows that *Am*-hsp83, in particular, has promising utility as a heat- or chill-inducible promoter; however, none of the known honey bee promoters would allow for antenna-specific expression.

While both expression cassettes have potential caveats (*i.e.* global expression and T2A residual tags), our future goal is to evaluate transgenic adults to see precisely what effect this has on honey bee development and physiology. If the potential limitations do not prohibit biologically realistic experiments, this system offers a rich future for further deciphering the roles that OBP16 and OBP18 play in disease odorant sensitivity and hygienic behaviour in general.

5.7 Methods

5.7.1 Ligand binding assays

Recombinant OBP16 and OBP18 were cloned, expressed, and purified exactly as previously described[106]. Briefly, the OBP genes were PCR amplified from honey bee cDNA and cloned into

a PET-5b bacterial expression vector. Plasmids were transformed into BL21(DE3)Rosetta-gami (OBP16) and BL21(DE3)pLysS *E. coli* strains and protein expression was induced via IPTG. The recombinant proteins were then purified by a series of chromatographic elutions, including anion exchange (DE-52, QFF, or Mono-Q) and gel filtration (Sephacryl-100 or Superose-12) as well as other standard purification protocols[258,262].

We then used an NPN (N-Phenyl-1-naphthylamine) competitive binding assay to measure relative affinities of β-ocimene, oleic acid, phenethyl acetate, and hexane (negative control). Binding assays were also conducted as previously described,[106] except they were performed in technical duplicate with 2 μM protein, 2 μM NPN, and between 0 and 8 μM of hexane and phenethyl acetate or between 0 and 6 μM of β-ocimene and oleic acid.

5.7.2 **dsRNA sequence selection, synthesis, and purification**

OBP16 and 18 have a large degree of sequence similarity to other honey bee OBPs, therefore, dsRNA sequences were selected by the following criteria: the region must not contain any 21 nucleotide stretch with more than 90% identity to another honey bee gene (*i.e.* it must have ≤ 19/21 nucleotide homology). The OBP genes are also very short (~400 nucleotides long), which further limits the options. For OBP16, the longest candidate dsRNA sequence which could also be amplified by RT-PCR without contamination of other sequences was 144 bp, while for OBP18 it was 115 bp. The control GFP sequence was 240 bp. These sequence lengths include the added 5' and 3' T7 polymerase promoter sequence to facilitate *in vitro* transcription of the dsRNA sequences (see Table 5.3 for primer sequences).

Table 5.2. Primer sequences for RT-PCR amplification of segments for dsRNA synthesis

Name	Sequence
OBP16_For	ACT GAT TAA TAC GAC TCA ACT ATA GGG GTT GGT GCA ATG ACA CAT GA
OBP16_Rev	ACT GAT TAA TAC GAC TCA ACT ATA GGG TTC ATC TAT TAT TTT TTG AC
OBP18_For	ACT GAT TAA TAC GAC TCA ACT ATA GGG CTA TCT CTG ATG CTG ACT TA
OBP18_Rev	ACT GAT TAA TAC GAC TCA ACT ATA GGG ATT TTC CAA TAC ACT TCA AT
GFP-For	AAT ACT CGA GTA ATA CGA CTC ACT ATA GGG AGC TGT TCA CCG GGG TGG
GFP-Rev	AAT ACT CGA GTA ATA CGA CTC ACT ATA GGG GTA GGT GGC ATC GCC CTC

Total RNA was extracted from honey bee worker antennae using Trizol reagent according to the manufacturer's protocol. cDNA was synthesized using SuperScript III reverse transcriptase according to the manufacturer's protocol and the desired OBP16 and OBP18 segments were subsequently PCR-amplified using the primers in Table 5.3. The PCR was analyzed by agarose gel electrophoresis, then the band was gel-purified (QIAquick gel extraction kit, Qiagen) and sequenced by Sanger sequencing (NAPS DNA sequencing facility, UBC) to confirm sequence purity and specificity. Bi-directional *in vitro* transcription was then performed using T7 polymerase (NEB) according to the manufacturer's protocol. The dsRNA was purified using an RNEasy kit (Qiagen), suspended in nuclease-free water, and quantified by nanodrop.

Since longer (>300 bp) dsRNA sequences are reported to be more efficacious[249], we later redesigned the dsRNA sequence to include repeats of the shorter, unique regions (Table 5.4). The GFP sequence remained the same as before, at 252 bp, but the OBP16 sequence was modified to include two repeats of a 219 bp unique region, which we were able to amplify above but not in sufficient purity to commence with the T7 reaction, separated by a random 7-mer. OBP18 was modified to include five repeats of a 66 bp unique region and one 102 bp unique

region, with repeats also separated by the same random 7-mer. These sequences were synthesized by AgroRNA (Seoul, South Korea) and purified by either their proprietary method or by isopropanol precipitation.

Table 5.3. Redesigned dsRNA sequences for synthesis by AgroRNA.

GFP dsRNA sequence
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACG TAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT GAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACG GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAG
OBP18 dsRNA sequence*
GTACAAT**TTAAACTATAAAAAGATCTTGTACATTCGACGTTTCTTGAGTTTTCCTTAAAAATATTT** **CAAGAAC**GTACAAT**TTATCTCTGATGCTGACTTAGCTGTAAAATCTGCTAAATTATTGAAGTGTAT** **TGGAAAAT**GTACAA**TTATCTCTGATGCTGACTTAGCTGTAAAATCTGCTAAATTATTGAAGTGTA** **TTGGAAAAT**GTACAAT**TTATCTCTGATGCTGACTTAGCTGTAAAATCTGCTAAATTATTGAAGTGT** **ATTGGAAAAT**GTACAAT**TTATCTCTGATGCTGACTTAGCTGTAAAATCTGCTAAATTATTGAAGTG** **TATTGGAAAAT**GTACAAT**GTAACACAATATTTTTTCTTTATTTTAAAATTGTTTTAATTATACTTTG** **ATTATAATTATATTAATTATACTTTATTATTATACTTTTAACTTTTATTATAT**GTACAAT
OBP16 dsRNA sequence*
GTACAAT**GCATATTCGATATTTGTTCAGTTCTCGTTGAACGTTTCAAGAATAGTCGAGTATTTTTA** **TTTATTTGAAATCGAtGTACAATGTTGGTGCAATGACACATGAGGAATTAAAAACCGGAATACAG** **ACTTTACAGCCAATTTGCGTAGGCGAAACTGGCACTAGTCAAAAAATAATAGATGAAGTTTATAA** **TGGCAACGTCAATGTAGAAGACGAAAATGT**GTACAAT**GTTGGTGCAATGACACATGAGGAATTA** **AAAACCGGAATACAGACTTTACAGCCAATTTGCGTAGGCGAAACTGGCACTAGTCAAAAAATAA** **TAGATGAAGTTTATAATGGCAACGTCAATGTAGAAGACGAAAATGTGTACAATTAAAATATATTT** **GAAAACTTTTATTAATAAATCAATATTATATATTATTATAAATTAATTATTA**GTACAAT

*Bold blue letters indicate regions matching to the OBP

5.7.3 **In-hive dsRNA larval feeding**

To obtain uniformly-aged brood, the queen was confined to a single empty frame using a one-frame queen excluder. After one day, the queen was released, and after five days, we retrieved the frame from the hive and dispensed 1 µl (2 µg/µl) of dsRNA into the droplet of brood food

bathing the larva at the base of the cell. We treated patches of ~20 cells (1 patch = 1 replicate) with either OBP16, OBP18, or GFP dsRNA (3 replicates per treatment). Patch locations were traced on a plastic transparency and when the cells were capped, the frame was retrieved from the hive and kept in a humid 33^{o}C incubator for the remainder of development. Bees were caught as they emerged, euthanized on dry ice and their antennae were dissected for protein analysis.

5.7.4 *In vitro* larval rearing and dsRNA larval feeding

Newly hatched larvae were grafted from a naturally laid frame and transferred to float in warmed brood food in 24-well plates. We used the basic larval diet composed of 53% lyophilized royal jelly, 3% glucose, 3% fructose, 1% yeast extract, and 40% water. Larvae were maintained in a humid incubator at 33^{o}C and were transferred to new plates with new food each day. Once uric acid crystals were visible, the larvae were transferred to a tissue-lined petri dish, where they remained for pupation and subsequent development. For the dsRNA treatments, the larvae were reared on *in vitro* transcribed dsRNA-containing brood food (50 µg/ml) for 3 d (for their 2nd – 5th instars).

5.7.5 Adult feeding

Newly emerged adult honey bees were collected from a single colony and maintained in plastic, ventilated cages (~20 bees per cage = 1 replicate) with excess 50% sucrose syrup fed through a punctured 15 mL Falcon tube. For the dsRNA treatments, we fed bees with either GFP dsRNA or a 1:1 w/w blend of redesigned OBP16 and OBP18 dsRNA at 50 µg/mL (5 replicates each). Bees were treated for 5 d before euthanizing them and dissecting their antennae for protein analysis.

5.7.6 **Antennal microinjection**

Pink-eyed pupae were carefully removed from their capped brood cells using forceps and placed in a tissue-lined petri dish (15 bees per dish = 1 replicate). Each antenna flagellum was injected near the joint between the flagellum and the scape with ~50 nL of solution containing 2 µg/µl of either OBP18 or GFP dsRNA (3 replicates each). We used home-made borosilicate glass injection needles and an Ependorf Femtojet 4i self-pressurized microinjector equipped with a MK1 manual micromanipulator (Singer Instruments). Pressure was set to 30 kPa, with a balance pressure of 5 kPa. Pupae were then maintained in a humid incubator at 33^{o}C until they reached adulthood, then were euthanized and their antennae were dissected for protein analysis.

5.7.7 **Protein extraction and processing**

Protein was extracted, digested and prepared for mass spectrometry exactly as described in Section 3.8.6, except stable isotope (SIS)-labelled peptides corresponding to OBP16 and OBP18 were added prior to digestion (Table 5.5). No SIS peptides were included for OBP21.

Table 5.4. Peptides targeted in multiple reaction monitoring experiments.

Peptide sequence[1]	Protein	fmol peptide/µg protein	Transitions[2]
EIAEIYLDENEVNK	OBP21		839.9(+2) → **960.5(+1),** 1123.5(+1), 1436.7(+1)
NGIIDVENEK	OBP21		565.8(+2) → 618.3(+1), **733.3(+1),** 846.4(+1)
TGIQTLQPICVGETGTSQK	OBP16		673.3(+3) → 638.8(+1), 742.4(+1), **807.4(+1)**
TGIQTLQPIC̲VGETGTSQ**K**[3]	OBP16	200	676.0(+3) → 642.8(+1), 742.4(+1), **815.4(+1)**
TITDILNS	OBP16		438.7(+2) → 771.4(+1), **657.4(+1),** 544.3(+1)
T**I**TDILNS[4]	OBP16	200	442.2(+2) → 778.4(+1), **664.4(+1),** 551.3(+1)
EIAEIFLDENGVNK	OBP18		795.9(+2) → 775.4(+1), **1035.5(+1),** 1148.6(+1)
EIAEIFLDENGVN**K**	OBP18	100	799.9(+2) → 783.4(+1), **1043.5(+1),** 1156.6(+1)
IETSIDQQK	OBP18		531.3(+2) → 518.3(+1), 718.4(+1), **819.4(+1)**
IETSIDQQ**K**	OBP18	100	535.3(+2) → 526.3(+1), 726.4(+1), **827.4(+1)**
DGNIDVEDEK	OBP18		567.3(+2) → 619.3(+1), **734.3(+1),** 847.4(+1)
DGNIDVEDE**K**	OBP18	100	571.3(+2) → 627.3(+1), **742.3(+1),** 855.4(+1)

1. Heavy residues within the peptide sequence are in bold and lines with heavy peptides are shaded grey
2. Only bold transitions were used for quantification
3. The underlined cysteine was reduced and alkylated during peptide synthesis
4. This is a C-terminal peptide; therefore, an internal residue (I) was labeled

5.7.8 **Mass spectrometry analysis**

Peptide samples (8 µg each, determined at the protein level by a BCA assay) were analyzed on an Agilent 6460 QQQ mass spectrometer essentially as previously described[112]. Monitored transitions can be found in Table 5.5. Peak areas for the quantifying transition were manually extracted and integrated using Mass Hunter Qualitative Analysis software (vB.06.00). The area under the target peptide's transition curve was normalized to the area under the SIS peptide transition's curve to yield comparable quantities across analyses. Normalized values were averaged across peptides for the same protein. Then, these values were further normalized to the average expression of that protein found in the GFP dsRNA-treated samples in order to express quantities as a fraction of the negative control. Statistics were performed using the Microsoft Excel Data Analysis Package (2010).

5.7.9 **Transgene cassette development and expression confirmation**

The piggyBac transgene vector backbone as well as the template plasmid for transposase synthesis were kindly supplied by Dr. Christina Schulte and Dr. Martin Beye. For vectors other than the original GFP-containing piggyBac cassette, inserts were designed *in silico* and synthesized by GenScript (NJ, USA), resulting in the architecture depicted in Figure 5.8.

To confirm that the insert contained all the proper genetic elements for expression, we transfected *D. melanogaster* S2 cells using XtremeGene (according to the manufacturer's protocol for Sf9 insect cells) and analyzed the cells for GFP expression by fluorescence microscopy and Western blot. Cells were incubated at 28ºC for 48 h post-transfection before fluorescence imaging or protein extraction. For the test for GFP dsRNA activity, transfected cells were then treated with dsRNA using the soaking method, as previously described[263]. Following this, cells were incubated overnight prior to protein extraction and analysis. Florescence microscopy images were taken using a Cellomics High-Content Screening microscope (Thermo) using the 485_BGS_BGS excitation/emission filter. Fluorescence microscopy was performed in a non-quantitative manner and was intended as an initial screen prior to Western blot analysis.

Protein was extracted from transfected or dsRNA-treated S2 cells and 20 µg was loaded in each lane of a precast gradient gel (4-16% polyacrylamide; Bio-rad). Protein was transferred to a nitrocellulose membrane via semi-dry transfer (50 mA, 30 min), blocked in 5% skim milk TBST for 30 min, then co-probed for GFP (27 kDa) and actin (42 kDa; loading control) before washing (5 x 5 min TBST), secondary incubation, washing again, and HRP imaging (Clarity™ Western ECL, Bio-rad). Primary antibody sources were DSHB (GFP-1D2) and Abcam (ab8224).

5.7.10 **Honey bee egg collection, microinjection, Western blot analysis, and imaging.**

Honey bee eggs were collected from colonies using the Jenter™ system, injected with ~400 pL of plasmid (75 ng/µl) and transposase (150 ng/µl), and maintained in an incubator exactly as described by Schulte *et al.*[250] Transposase was synthesized and purified exactly as previously described using mMessage mMachine T7 synthesis kit and the MegaClear purification kit (Ambion). After 2 d, eggs that remained intact were harvested and protein was extracted following the same protocol as for antennae (Section 3.8.6). 100 µg of protein was loaded in each lane for the Western blot; otherwise, the protocol was performed as described in Section 5.7.8. Once hatched, some of the larvae were analyzed by fluorescence imaging as described in Section 5.7.8, except all images were taken at 5x magnification. Live larvae (*i.e.* they were visibly moving under a dissecting microscope) were carefully transferred to a 96-well imaging plate using a soft paintbrush and imaged immediately.

Chapter 6: Discussion

Over the last decade, honey bee research has been a burgeoning field. Owing to bees' essential role in agricultural food production, a keen stakeholder and public interest, and the rise of modern 'omics technologies, we now know more than ever before about honey bee host-pathogen and host-parasite interactions. Despite the advances in honey bee and *Varroa* genome sequencing, this information cannot yet be fully utilized because their databases of predicted genes and proteins need further refinement to cover missing sequences. In addition, although there have been many differential expression studies investigating honey bee disease defenses, such as hygienic behaviour, there has been very little work on the underlying molecular mechanism. Some work has been done to identify hygienic behaviour-inducing odorants, with the most in-depth analysis being conducted by Swanson *et al.*[94] on odorants emitted from chalkbrood-infected larvae. However, odorants from FKB have never been investigated, despite being an alluring model system for studying hygienic behaviour. Identifying more odorants and odorant blends that elicit the behaviour, and their interactions with hygienic behaviour-associated proteins, is a gap in our understanding of how this behaviour works.

In an effort to fill these gaps, we set out asking several questions: 1) Could there be novel protein-coding regions in the honey bee and *Varroa* genome and 2) Can we construct new protein databases to learn more about these organisms' fundamental biology (both addressed in Chapter 2)? We followed this by investigating some of the molecular mechanisms underlying host-parasite and host-pathogen interactions in honey bees: 3) Can we identify specific hygienic-behaviour stimulating odorants associated with freeze-killed and *Varroa*-infested brood that are absent or reduced in healthy brood (Addressed in Chapters 3 and 4)? Finally, we asked: Do these

odorants interact with hygienic behaviour-associated odorant binding proteins (addressed in Chapter 5)? We have successfully answered all four questions, despite not all our approaches being effective. Here we will discuss the strengths and limitations of our work in the broader context.

6.1 Proteogenomics: Insights, limitations, and implications

In proteomics, we can typically only identify proteins if their amino acid sequences are already known. Therefore, having a comprehensive protein database underpins our ability to identify changes in protein expression that define biological processes. Arguably, we have not yet fully defined the proteome of any organism, with the most well-developed proteomes being for only the most-studied organisms. The genomic research community for honey bees and *Varroa* are relatively small so the current proteome databases for these organisms are still relatively crude. Therefore, improving these databases will allow us to better interrogate samples for proteomic shifts.

Our *Varroa* and honey bee proteogenomics investigations were able to define many candidate novel protein-coding regions, which supports our hypothesis that there could be unannotated genes in these species. Further support for this came by confirming that many of the candidate protein-coding regions had significant sequence similarity to annotated regions in other arthropods, showing that some of the regions are located very close to one another in the genome (resembling exons), and in the case of the honey bee sequences, many overlap with existing RNA-seq data and some were confirmed by RT-PCR. However, this is still a preliminary investigation into ORF expression and possible function. More information will need to be

gathered before the provisional ORF sequences we identified can be integrated into gene models, for example, by mining RNA-seq data to identify exons, termini, and splice sites not covered by mass spectrometry data.

Based on the information we have obtained, we are unable to say what the complete gene sequences indicated by these candidate novel ORFs are. Rather, we see the main strength of this work as being that we have generated new peptide expression data that can be integrated into gene prediction algorithms to help better-train the algorithm to identify genes (*i.e.*, make the algorithm more sensitive). In addition, since we have integrated the candidate sequences into fasta databases, we can continue to identify these sequences in proteomics experiments and gain more insight into their biological relevance while improving coverage. We think that using even provisional expression information is appropriate for this purpose, especially for non-model species which have not had thoroughly refined annotations (as opposed to humans or other more thoroughly refined species, which are expected to have far fewer missing or mis-annotated genes). We cannot afford to wait for annotation refinements which may or may not occur. Although imperfect, our approach at least enables us to catalog a more comprehensive proteome, furthering our proteomics capabilities in the meantime.

While there are bound to be some false positive peptide and protein sequences in our data, we expect that the data will still allow us to decrease false negatives through more accurate gene predictions. Indeed, other lines of evidence typically integrated into gene prediction algorithms include gene sets from entirely different species, many of which are also false positives with respect to the target genome. However, prediction algorithms also look for other genetic

signatures and use homology information during the annotation process. Therefore, at this stage, we think it is beneficial to be as inclusive as possible. If the annotation algorithm predicts new genes based on this expression data, each novel gene prediction can then be investigated in detail, as output by the algorithm.

Since *Varroa* has only had one concerted annotation effort, our proteogenomics investigation also provided an opportunity to look into why the candidate novel protein-coding regions we identified were missed in the previous annotation. For honey bees, one reason why previous annotations suffered from low-accuracy is because there is substantially divergent codon usage in the honey bee genome[140]. That is, there is a bias towards using redundant codons with an A or T in the 3' position, rather than G or C. Surprisingly, in *Varroa*, we were unable to identify any differences in nucleotide or codon composition between the novel and known regions. Therefore, our data do not support the idea that annotations were missed because of atypical genomic composition. While the reason behind the failure for these sequences to be annotated remains unclear, we still see several avenues forward for improving the *Varroa* annotation in the future.

Since there have been very few 'omics expression analyses of *Varroa* (transcriptomic or proteomic), this is one clear way to help refine the annotation. Generating libraries of transcriptomic data from *Varroa* and other mite species may offer a better informant gene set and allow more candidate gene sequences to be identified. To our knowledge, this has not yet been conducted. Our own proteogenomics peptide dataset can be included as informant data as well, and as gene annotation software becomes more modernized to better-include peptide and mass spectrometry data inputs in general, we expect that these challenges with annotating non-model

species will diminish over time. In the meantime, we have shown through the large number of candidate novel protein-coding regions – indeed, some of which were even significantly differentially expressed different developmental stages – that *Varroa* research could benefit substantially from a refinement effort.

Honey bees, however, have already undergone several iterations of annotation refinement. Nevertheless, our data suggest that there is still considerable improvement needed. Thousands of sequences were not retained from OGSv1.0 to OGSv3.2 because they were not supported by empirical evidence (transcripts) at the time, but when MS/MS data was searched against a composite database, 8% of all identified peptides (representing 746 proteins) matched to sequences found only in OGSv1.0, indicating that many proteins (or fragments of the protein containing these peptides) may have been erroneously dropped in OGSv3.2. This is therefore a cautionary tale of where annotation refinement process can actually introduce additional errors in the database.

While the number of newly identified candidate coding regions in honey bees is substantial, it is not enough to explain the discrepancy we observe between peptide identification rates in honey bees and other commonly studied organisms. Previously, we also investigated other explanations for missed identifications (*i.e.* unexpectedly high genetic diversity in honey bees, post-translational modifications causing mass shifts that were unaccounted for, high endogenous protease activity, and potential limitations in dynamic range of the mass spectrometer)[151]. However, our data did not support that any of these alternate explanations were substantial contributors to missed identifications either. Despite identifying fewer novel protein-coding

regions than we hoped in the present work, we still suspect that poor annotation quality relative to model species is the main contributing factor. This is because while proteogenomics allows us to survey the genome for expression of candidate unannotated peptides, it is still a relatively insensitive method owing to escalating false discovery rates with database size; therefore, we expect that there could be many more unannotated regions than the ones we were able to identify here.

6.1.1 **Functional annotations**

Whether new gene sequence annotations come to fruition or not, assigning functions to the genes that have been identified thus far is a future direction of utmost importance. Annotation pipelines typically work well among vertebrates; however, invertebrates generally have high rates of sequence diversity and longer evolutionary distances[264]. Orthology delineation of genomes from ten bee species showed that ~1,200 to ~2,800 genes in each species (10%-21%) had no detectable orthology beyond Hymenoptera[265] (*i.e.*, they are taxonomically-restricted genes, or TRGs), preventing any functional prediction based on orthologs alone. Furthermore, in honey bees, only 52% of genes in OGSv3.2 are linked to GO terms, meaning that only about half of differentially expressed genes may contribute to interpreting differential expression analyses via, *e.g.*, functional enrichment tests.

‘Omics technologies open the possibility of hypothesis-free approaches to finding new candidate protein-coding regions without any *a priori* knowledge. However, by applying automatic functional prediction and not pursuing genes with unknown function, proteomics and transcriptomics experiments do not reach their true potential. Once new candidates are identified,

such as those in our proteogenomics effort, we hope that targeted gene manipulation combined with global expression studies and improved gene databases will aid with unravelling gene functions. Until then, we will continue to utilize these new provisional protein databases in our future proteomics experiments to investigate hygienic behaviour and other host-parasite interactions.

6.2 Odorant identification and behavioural assays: Insights, limitations and future directions

Hygienic behaviour is known to depend on worker bees' olfactory sensitivity to disease odorants, but very few of these odorants have been identified and confirmed to be functional in behavioural assays to date. In this section of the thesis, we asked the question: Can we identify specific hygienic-behaviour stimulating odorants emitted from FKB and *Varroa*-infested brood that are absent or reduced in healthy brood? We were able to successfully identify candidate hygienic behaviour-inducing odorants from FKB and validate two of these candidates (β-ocimene and oleic acid) behaviourally. Our data therefore support our initial hypothesis that specific hygienic behaviour-inducing odorant cues are released from FKB; however, our data do not support this hypothesis for *Varroa*-infested brood.

We did not find specific odorants that were differentially emitted from brood that had been parasitized at any of the developmental stages we tested (white-eyed pupae through purple-eyed, tan body pupae). It is possible that the odorants either 1) exist below our GC-MS limit of detection, 2) they have chemical properties that are incompatible with our extraction methods, or 3) they do not exist in our particular samples. This third point may be the most important of them

all; previous research has identified a brood effect contributing to *Varroa*-resistance in Eastern honey bees[30], as well as VSH in Western honey bees[85,89]. A proposed mechanism of this brood effect is that *Varroa*-parasitized brood exhibiting this effect likely emit stronger odorant signals than their Western honey bee or low-VSH counterparts. Our samples, however, came from Western honey bee colonies with a wide range of hygienicity, but have not been tested for VSH. Therefore, they may emit only a weak odorant signal, or no signal at all. Future work to concentrate the odorant samples, explore different extraction approaches, and utilize different sample sources (*i.e.* from Eastern honey bee brood or high VSH Western honey bee brood) will be necessary to identify putative hygienic behaviour-inducing compounds from *Varroa*-infested brood.

6.2.1 **FKB as a model system**

Since the 1990s, there has been an ongoing debate over whether hygienic colonies are also resistant to *Varroa*[266]. Some experiments suggest that the relationship between hygienic behaviour and *Varroa* resistance is weak[267], while others find that it is strong and significant[87,112,268]. Nevertheless, there is consensus in the field that both traits rely at least in part on the bees having high olfactory sensitivity, and that VSH also depends on an added brood effect[89]. One such brood effect was proposed by Page *et al.*[30], who observed that Eastern honey bee brood tend to self-destruct when infested with *Varroa*. If this "social apoptosis" mechanism holds true for high VSH Western honey bees, a unifying hypothesis to test in the future is that odorants associated with death may actually be the same ones that enable VSH via social apoptosis. Perhaps hygienic behaviour and VSH are not distinguished by olfactory sensitivity to different odorants, but whether or not the brood self-sacrifices upon mite infestation to produce a

detectible odorant signal. This could explain why some hygienic colonies, but not others, are also resistant to mites – the resistant colonies could be the ones whose brood emit a stronger signal. Indeed, Hu *et al.*[115] performed a differential proteomics study comparing antennae of honey bees from high and low VSH colonies and found that one of our very own markers for hygienic behaviour – OBP18 – was among those that were upregulated in high VSH bees. Parker *et al.*[38] compared similar samples, including both VSH and hygienic behaviour comparisons, but did not identify OBP18; however, their proteomics dataset was not as deep as Hu *et al.* Nevertheless, the confirmed presence of up-regulated OBP18 in at least some hygienic and VSH populations suggests that these bees could be detecting at least some of the same odorants.

FKB could be a fruitful model system for not only social apoptosis, but also for studying social immunity across insect species (namely, ants, termites, wasps, and bees). Different insect species experience a wide range of pests and pathogens, which likely also trigger emission of different necrophoresis-inducing odorants; therefore, it would be difficult if not impossible to directly compare disease-response mechanics between species. However, death is common to all of them, and is the ultimate outcome of most of the brood diseases. Since we identified one evolutionarily conserved necromone in our honey bee FKB analysis (oleic acid), we think this could be an appropriate method to investigate parallel social immunity mechanisms across other species, too. Freeze-killing may not be a typical means of death that honey bees (or other social insects) experience, but we know from years of selective breeding programs that colonies that quickly detect and remove FKB also quickly detect and remove diseased brood. Our identification and behavioural validation of oleic acid as a hygienic behaviour-inducing odorant in honey bees

leads us to postulate that adaptations to the FKB assay and the odorants emitted from FKB could be used more broadly to investigate evolutionary parallels of social immunity.

6.2.2 **Limitations of electroantennography**

The electroanennography investigations conducted here were not as informative as we had expected. Although we initially identified an intriguing lateralized antennal response to β-ocimene (*i.e.* the left antenna of hygienic bees responded more strongly than the right, but both sides responded in the same way for non-hygienic bees), we found that upon closer inspection, this was actually a result of side-dependent responses to the solvent, not the odorant. We are still not sure why the solvent induced a lateralized response, but because of this we can conclude that there is no association between hygienic behaviour and lateralization. As such, the only information we really gain from the EAG experiments is whether or not honey bees have the potential to sense the candidate odorants at all (*i.e.* whether they induce dose-dependent antennal nerve responses). Even there, the results have limited utility, as evidenced by the apparent lack of stimulation by oleic acid, but significant behavioural response even when applied beneath an intact cell cap. EAG may still be a useful method for testing how targeted gene manipulation may affect antennal sensitivity, but based on our data, we do not think it is an optimal tool for screening candidate hygienic behaviour-inducing odorants. The front-way odorant assay, on the other hand, is actually faster than EAG, and likely a better screening assay.

6.2.3 **Combined behavioural assays yield mechanistic insights**

We used a combination of three different behavioural assays to evaluate the effectiveness of FKB odorants to induce hygienic behaviour in the hive setting (the front-way odorant assay,

back-way odorant assay, and broodless back-way odorant assay). By considering data from these three approaches, we are beginning to disentangle some of the complex outcomes we have observed.

While the results of our front-way odorant assays do not account for the effect of new odorants that could be emitted as brood respond to the odorant application, it still offers a quick screening method for candidates. The best candidates can then be investigated further, ideally using an assay that 1) allows for full uncapping and removal behaviour, 2) requires permeation through an in-tact cell cap, and 3) involves no confounding brood response to the odorant. The back-way odorant assays (which were conducted with and without brood) achieve all three of these criteria. Swanson *et al.*[94] used a brood dummy odorant impregnation assay, in which parafin wax dummies were treated with candidate odorants and then bees were observed to either cap the cell (indicating acceptance) or not (indicating 'hygienic behaviour'). This assay is an elegant approach, but while it achieves criteria 2) and 3), it does not achieve 1). That is why we developed a variation of the backway odorant assay in which we remove the developing brood through the back of a Jenter cage, then insert an odorant-treated object and re-sealed the cell. Importantly, Swanson *et al.* observed that when methyl linolenate was incorporated into their wax dummies, bees capped the cells as if they were 5^{th} instar larvae. Methyl linolenate is, therefore, a good candidate hygienic behaviour suppressor that we should investigate with broodless back-way assays in the future. The presence of this compound should inhibit the uncapping of empty and hexane-treated cells, allowing us to better-test the ability of odorants to induce hygienic behaviour in the absence of brood. This would be the most controlled measure

of odorant-induced hygienic behaviour while maintaining as biologically realistic of a scenario as possible.

6.2.4 **Odorant context**

Cumulatively, the results of our behavioural assays suggest that the context in which bees encounter an odorant can be important in dictating the behaviour that ensues. For example, in the front-way odorant assays, oleic acid, β-ocimene, and a blend of the two all induced high levels of hygienic behaviour. Presumably this is because the odorants were applied to brood heads and left uncapped, making the odorant signal immediately detectable and strong relative to the background. However, in the back-way odorant assays, β-ocimene did not have as strong of an effect. In particular, it did not induce hygienic behaviour when applied to prepupae, and while it tended to induce more hygienic behaviour when applied to pupae, the difference was not significant. This could be a combined effect of prepupae normally emitting higher levels of β-ocimene than pupae, in addition to the odorants perceived through the cap of the back-way assays likely being much weaker signals than in the front-way assays. Therefore, the application of β-ocimene to prepupae (which already emit the odorant) tended to be perceived as a more normal state than when it was applied to pupae (which normally emit less of the odorant). Oleic acid, which is not normally emitted by live brood, tended not to be perceive as normal in either instance. Interestingly, in the broodless back-way assays, β-ocimene had a suppressive, rather than stimulating, effect on hygienic behaviour (even when oleic acid was also present), while hexane-treated broodless cells were uncapped most frequently. In other words, cells that were not treated with β-ocimene or the blend were perceived as the most abnormal, even though cells treated with the blend induced high levels of hygienic behaviour before. Therefore, we think that

the death or disease odorant itself is not the only factor governing a hygienic response, but also the context in which it is encountered.

One hypothesis to test in the future is that hygienic behaviour may also be triggered by how "out of place" a brood cell appears. That is, how does the odorant profile compare to the surrounding cells? For example, a broodless, hexane-treated capped cell may be perceived as more abnormal when it is surrounded by normally developing brood (as is the case in the back-way assay) than when it is surrounded by empty, uncapped cells. Likewise, a broodless, brood pheromone-treated capped cell may be perceived as more similar to surrounding brood cells than an empty cell. If supported, this hypothesis could help explain why brood cells which are developmentally delayed compared to their immediate neighbours are targeted more often in VSH colonies[96]. An odorant profile associated with delayed development (*i.e.* a "younger" profile) may not be abnormal on its own, but when surrounded by older cells, it is.

6.3 Odorant interactions with OBP16 and OBP18

Odorant binding proteins have been positively correlated with hygienic behaviour and VSH in multiple independent studies[39,106,112,115]. However, at the time of writing, no one has conducted targeted gene manipulations to test if these proteins are necessary or sufficient to enable disease odorant detection, nor have binding assays investigating candidate hygienic behaviour-inducing odorants (other than oleic acid) been conducted. Therefore, in this section of the thesis, we aimed to answer the question: Do death or disease odorants interact with hygienic behaviour-associated odorant binding proteins?

Determining whether hygienic behaviour-inducing odorants interact with OBP16 and OBP18 was the most challenging aspect of this thesis. We hypothesized that they are indeed interacting and aimed to test this using *in vitro* and *in vivo* techniques; however, we only succeeded in testing this *in vitro.* Ligand binding assays using recombinant OBP16 and OBP18 showed that β-ocimene and oleic acid both interact with at least one of OBP16 and OBP18, while hexane and phenethyl acetate do not strongly interact with either OBP. Importantly, this reflects our behavioural observation that hexane and phenethyl acetate induced low levels of hygienic behaviour, whereas β-ocimene and oleic acid induced moderate to high levels, depending on the behavioural assay. While these *in vitro* assays provide evidence that the ligands have the potential to physically interact with the OBPs, the physical environment of the ligand binding assay is very different from the sensillium lymph. Ligand binding assays, for example, have no opportunity for competition between ligands other than NPN and the odorant of interest. We are unfortunately lacking the biological context for this result, which RNAi and transgenics would allow us to achieve.

Given our exhaustive RNAi attempts, we think that the current manifestations of this method are unlikely to be feasible for knocking down antennal genes. RNAi is an undeniably useful method; however, not all tissues and genes are good targets and we clearly were not able to use it to answer the question we set out to. Our failure to do so is likely a result of both a difficult tissue for dsRNA delivery and having short gene targets with high sequence similarities to each other and other OBPs. Inserting a construct into the genome that enables constitutive expression of OBP-targeting shRNA may be a viable option; however, we think that complete knock-out via CRISPR technology may be the most promising alternative. Indeed, very similar tools and

instrumentation could be used to develop a CRISPR bee as we have set up for the piggyBac-derived transgenic system, but no one has yet created CRISPR honey bees. Another alternative strategy is to produce and purify more recombinant OBP16 and OBP18 in order to screen for inhibitory ligands. If an irreversible OBP16- and OBP18-specific inhibitor can be identified and delivered *in vivo*, this could effectively knock-out OBP function without needing to change expression levels.

We have unfortunately not reached the point where we can use the transgenic technology to test for a biological effect of increasing OBP expression; however, we have made substantial progress in getting the technique underway. With significant assistance from Dr. Beye's group in Germany, we have acquired all the tools and instrumentation necessary, and have successfully produced GFP-expressing larvae. While this on its own is not novel, we have demonstrated that the T2A cleavage system is likely a viable option for other honey bee expression studies, although more tests will be needed before we can assert its efficiency.

Although we did not produce adult transgenic honey bees, we think this is still a feasible technique that warrants further pursuit. Even confirming heterologous gene expression at the larval stage is a promising result, as it indicates that at the very least, we can use the technique to study effects of gene expression in early life stages, *e.g.*, mechanisms of embryogenesis. We have yet to confirm that the transgene was actually incorporated into the genome as opposed to being transiently expressed from the plasmid, but even transient expression of cassettes in early life stages is a powerful molecular tool.

One immediate future direction, which is currently underway, is to rear the larvae hatching from injected eggs *in vitro* and analyze them for genomic incorporation (by PCR) as well as OBP18 expression (by mass spectrometry). Normally, larvae do not produce OBP18; therefore, larvae with successful cassette incorporation are expected to have significantly higher OBP18 expression than wild-type larvae. We currently possess approximately 100 potentially transgenic larvae to screen. A second on-going future direction is to rear the larvae into queens, then screen the queens' drone offspring to measure germ-line transgene incorporation efficiency. If any queens produce drones carrying the transgene, the drones could either be used directly for EAG tests or sperm could be harvested from them to propagate the transgenic lineage. Although drones do not do hygienic behaviour, their antennae may still exhibit differential EAG responses if they are overexpressing OBP18. We currently have eight such laying queens, from which we have harvested 160 larvae to screen.

As one of extremely few laboratories in the world to even attempt this technique, we inevitably have much more work to do before it is streamlined for common use. Having a simpler biological system to apply it to (such as the development of embryos or young larvae), although not relevant to hygienic behaviour, may be a more fruitful use of the technique in the meantime. Once it is more robust, the potential applications will be endless.

Indeed, the reason for including multiple approaches to testing this hypothesis was because *in vivo* gene expression manipulation techniques are generally not well-developed for honey bees and we anticipated a high likelihood of failure. Nevertheless, due to our three-pronged approach (RNAi, transgenics, and binding assays), we were able to demonstrate relatively strong physical

OBP-odorant interactions with *in vitro* experiments, which mirror our behavioural data from the front-way odorant assay surprisingly well. Although the binding assays were *in vitro,* it offers some of the first mechanistic evidence of OBPs' involvement in detecting hygienic behaviour-inducing odorants, strengthening past correlational data.

6.4 Conclusion

Honey bee colonies are valuable model systems for studying social behaviour and are indispensable assets for agricultural operations. Since *Varroa* is now the most destructive honey bee pest on a global scale, understanding the mechanisms that govern *Varroa*-bee interactions will be crucial for preserving colony health in the future.

In this body of work, we began by conducting a survey of unannotated protein-coding regions within both *Varroa* and honey bees to produce more comprehensive protein databases for mass spectrometry-based proteomics and suggest that the peptide expression data should be used for annotation improvements. We also found evidence for resurrecting hundreds of previously annotated honey bee proteins which have been since discarded without clear evidence that they were incorrect. In the process, we gained significant insights into proteomic shifts that occur during *Varroa* development and sex differentiation that had not been previously documented. These protein databases are now openly accessible for anyone to use in future proteomics research on these organisms.

Next, using FKB as a model system for studying the molecular mechanism of hygienic behaviour, we identified several odorant molecules associated with death and investigated two of

them behaviourally. Our pinnacle goal was to investigate if two odorant binding proteins – OBP16 and OBP18 – are necessary and sufficient to enable hygienic behaviour, and while our attempts to investigate this were either unsuccessful (RNAi) or have not yet yielded finalized results (transgenics), our ligand binding assays show striking parallels between the strength of odorant interactions with these OBPs and the odorants' abilities to induce hygienic behaviour *in vivo.* While a large body of work has previously determined that hygienic bees have heightened olfactory sensitivity, these results are some of the first steps that have been taken to examine how hygienic behaviour depends on the expression of specific genes. There are many more pieces to fit in this mechanistic puzzle, and we hope that this work may serve as a foundation for future interrogation.

References

1 Readicker-Henderson, E. *A short history of the honey bee: Humans, flowers, and bees in the eternal chase for honey*. (Timber Press, 2009).

2 Ruttner, F. *Biogeography and taxonomy of honeybees*. (Springer Science & Business Media, 2013).

3 vanEngelsdorp, D. & Meixner, M. D. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology* **103, Supplement**, S80-S95, doi:http://dx.doi.org/10.1016/j.jip.2009.06.011 (2010).

4 Southwick, E. E. & Southwick, J. L. Estimating the Economic Value of Honey Bees (Hymenoptera: Apidae) as Agricultural Pollinators in the United States. *Journal of Economic Entomology* **85**, 621-633, doi:10.1093/jee/85.3.621 (1992).

5 Morse, R. A. & Calderone, N. W. The value of honey bees as pollinators of US crops in 2000. *Bee culture* **128**, 1-15 (2000).

6 Mukezangango, J. & Page, S. (ed and Agriculture and Agri-Food Canada Horticulture and Cross Sectoral Division) (Government of Canada, Canada, 2017).

7 Graystock, P., Blane, E. J., McFrederick, Q. S., Goulson, D. & Hughes, W. O. H. Do managed bees drive parasite spread and emergence in wild bees? *Int J Parasitol Parasites Wildl* **5**, 64-75, doi:10.1016/j.ijppaw.2015.10.001 (2016).

8 Potts, S. G. *et al.* Global pollinator declines: trends, impacts and drivers. *Trends Ecol Evol* **25**, 345-353, doi:10.1016/j.tree.2010.01.007 (2010).

9 Elsik, C. G. *et al.* Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* **15**, 86, doi:10.1186/1471-2164-15-86 (2014).

10 Consortium, H. G. S. Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* **443**, 931-949 (2006).

11 Von Frisch, K. The dance language and orientation of bees. (1967).

12 Wenner, A. M. & Johnson, D. L. Honeybees: do they use direction and distance information provided by their dancers? *Science* **158**, 1072-1077 (1967).

13 Neumann, P. & Carreck, N. L. Honey bee colony losses. *Journal of Apicultural Research.* **49**, 1-6 (2010).

14 Currie, R. W., Pernal, S. F. & Guzmán-Novoa, E. Honey bee colony losses in Canada. *Journal of Apicultural Research* **49**, 104-106 (2010).

15 Potts, S. G. *et al.* Safeguarding pollinators and their values to human well-being. *Nature* **540**, 220-229, doi:10.1038/nature20588 (2016).

16 Moritz, R. F. A. & Erler, S. Lost colonies found in a data mine: Global honey trade but not pests or pesticides as a major cause of regional honeybee colony declines. *Agriculture, Ecosystems & Environment* **216**, 44-50, doi:https://doi.org/10.1016/j.agee.2015.09.027 (2016).

17 Dietemann, V. *et al.* Varroa destructor: research avenues towards sustainable control. *Journal of Apicultural Research* **51**, 125-132 (2012).

18 Francis, R. M., Nielsen, S. L. & Kryger, P. Varroa-virus interaction in collapsing honey bee colonies. *PLoS One* **8**, e57540, doi:10.1371/journal.pone.0057540 (2013).

19 Guzman-Novoa, E. *et al.* Varroa destructor is the main culprit for the death and reduced populations of overwintered honey bee (Apis mellifera) colonies in Ontario, Canada. *Apidologie* **41**, 443-450, doi:10.1051/apido/2009076 (2010).
20 Le Conte, Y., Ellis, M. & Ritter, W. Varroa mites and honey bee health: can Varroa explain part of the colony losses? *Apidologie* **41**, 353-363 (2010).
21 Cornman, S. R. *et al.* Genomic survey of the ectoparasitic mite Varroa destructor, a major pest of the honey bee Apis mellifera. *BMC Genomics* **11**, 602, doi:10.1186/1471-2164-11-602 (2010).
22 Rosenkranz, P., Aumeier, P. & Ziegelmann, B. Biology and control of Varroa destructor. *Journal of invertebrate pathology* **103**, S96-S119 (2010).
23 Sumpter, D. J. & Martin, S. J. The dynamics of virus epidemics in Varroa-infested honey bee colonies. *Journal of Animal Ecology* **73**, 51-63 (2004).
24 Bernardi, S. & Venturino, E. Viral epidemiology of the adult Apis Mellifera infested by the Varroa destructor mite. *Heliyon* **2**, e00101, doi:10.1016/j.heliyon.2016.e00101 (2016).
25 Di Prisco, G. *et al.* A mutualistic symbiosis between a parasitic mite and a pathogenic virus undermines honey bee immunity and health. *Proc Natl Acad Sci U S A* **113**, 3203-3208, doi:10.1073/pnas.1523515113 (2016).
26 Di Prisco, G. *et al.* Varroa destructor is an effective vector of Israeli acute paralysis virus in the honeybee, Apis mellifera. *J Gen Virol* **92**, 151-155, doi:10.1099/vir.0.023853-0 (2011).
27 Rath, W. Co-adaptation of Apis cerana Fabr. and Varroa jacobsoni Oud. *Apidologie* **30**, 97-110 (1999).
28 Lin, Z. *et al.* Go East for Better Honey Bee Health: Apis cerana Is Faster at Hygienic Behavior than A. mellifera. *PLoS One* **11**, e0162647, doi:10.1371/journal.pone.0162647 (2016).
29 Boot, W. J. *et al.* Natural selection of Varroa jacobsoni explains the different reproductive strategies in colonies of Apis cerana and Apis mellifera. *Experimental and Applied Acarology* **23**, 133-144 (1999).
30 Page, P. *et al.* Social apoptosis in honey bee superorganisms. *Sci Rep* **6**, 27210, doi:10.1038/srep27210 (2016).
31 Fries, I., Imdorf, A. & Rosenkranz, P. Survival of mite infested (Varroa destructor) honey bee (Apis mellifera) colonies in a Nordic climate. *Apidologie* **37**, 564-570 (2006).
32 Korpela, S., Aarhus, A., Fries, I. & Hansen, H. Varroa jacobsoni Oud. in cold climates: population growth, winter mortality and influence on the survival of honey bee colonies. *Journal of Apicultural Research* **31**, 157-164, doi:10.1080/00218839.1992.11101278 (1992).
33 Currie, R. & Gatien, P. Timing acaricide treatments to prevent Varroa destructor (Acari: Varroidae) from causing economic damage to honey bee colonies. *The Canadian Entomologist* **138**, 238-252 (2006).
34 Pettis, J. S. A scientific note on Varroa destructor resistance to coumaphos in the United States. *Apidologie* **35**, 91-92 (2004).
35 Mikheyev, A. S., Tin, M. M., Arora, J. & Seeley, T. D. Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. *Nat Commun* **6**, 7991, doi:10.1038/ncomms8991 (2015).

36 Bixby, M. *et al.* A Bio-Economic Case Study of Canadian Honey Bee (Hymenoptera: Apidae) Colonies: Marker-Assisted Selection (MAS) in Queen Breeding Affects Beekeeper Profits. *J Econ Entomol* **110**, 816-825, doi:10.1093/jee/tox077 (2017).
37 Guarna, M. M. *et al.* Expression biomarkers used for the selective breeding of complex polygenic traits. *BioRxiv*, http://dx.doi.org/10.1101/076174 (2016).
38 Parker, R. *et al.* Correlation of proteome-wide changes with social immunity behaviors provides insight into resistance to the parasitic mite, Varroa destructor, in the honey bee (Apis mellifera). *Genome Biol* **13**, R81, doi:10.1186/gb-2012-13-9-r81 (2012).
39 Mondet, F. *et al.* Antennae hold a key to Varroa-sensitive hygiene behaviour in honey bees. *Scientific reports* **5**, 10454 (2015).
40 Locke, B. Natural Varroa mite-surviving Apis mellifera honeybee populations. *Apidologie* **47**, 467-482 (2016).
41 Panziera, D., van Langevelde, F. & Blacquière, T. Varroa sensitive hygiene contributes to naturally selected varroa resistance in honey bees. *Journal of Apicultural Research* **56**, 635-642, doi:10.1080/00218839.2017.1351860 (2017).
42 Smith, K. M. *et al.* Pathogens, pests, and economics: drivers of honey bee colony declines and losses. *Ecohealth* **10**, 434-445, doi:10.1007/s10393-013-0870-2 (2013).
43 Chan, Q. W., Melathopoulos, A. P., Pernal, S. F. & Foster, L. J. The innate immune and systemic response in honey bees to a bacterial pathogen, Paenibacillus larvae. *BMC Genomics* **10**, 387, doi:10.1186/1471-2164-10-387 (2009).
44 Wilson, W. T. Resistance to American foulbrood in honey bees: XII. Persistence of viable Bacillus larvae spores in the feces of adults permitted flight. *Journal of Invertebrate Pathology* **20**, 165-169 (1972).
45 Hansen, H. & Brødsgaard, C. J. American foulbrood: a review of its biology, diagnosis and control. *Bee World* **80**, 5-23, doi:10.1080/0005772X.1999.11099415 (1999).
46 Kochansky, J., Knox, D. A., Feldlaufer, M. & Pettis, J. S. Screening alternative antibiotics against oxytetracycline-susceptible and-resistant Paenibacillus larvae. *Apidologie* **32**, 215-222 (2001).
47 Spivak, M. & Downey, D. L. Field assays for hygienic behavior in honey bees (Hymenoptera: Apidae). *Journal of economic entomology* **91**, 64-70 (1998).
48 Spivak, M. & Gilliam, M. Hygienic behaviour of honey bees and its application for control of brood diseases and Varroa: Part II. Studies on hygienic behaviour since the Rothenbuhler era. *Bee world* **79**, 169-186 (1998).
49 Spivak, M. & Gilliam, M. Hygienic behaviour of honey bees and its application for control of brood diseases and varroa: Part I. Hygienic behaviour and resistance to American foulbrood. *Bee World* **79**, 124-134 (1998).
50 Spivak, M. & Reuter, G. Resistance to American foulbrood disease by honey bee colonies Apis mellifera bred for hygienic behavior. *Apidologie* **32**, 555-565 (2001).
51 Aronstein, K. & Murray, K. Chalkbrood disease in honey bees. *Journal of invertebrate pathology* **103**, S20-S29 (2010).
52 Gilliam, M., Taber III, S. & Richardson, G. V. Hygienic behavior of honey bees in relation to chalkbrood disease. *Apidologie* **14**, 29-39 (1983).
53 Palacio, M. A., Rodriguez, E., Goncalves, L., Bedascarrasbure, E. & Spivak, M. Hygienic behaviors of honey bees in response to brood experimentally pin-killed or infected with Ascosphaera apis. *Apidologie* **41**, 602-612 (2010).

54 Liu, Y. *et al.* Larva-mediated chalkbrood resistance-associated single nucleotide polymorphism markers in the honey bee Apis mellifera. *Insect Mol Biol* **25**, 239-250, doi:10.1111/imb.12216 (2016).
55 Galbraith, D. A. *et al.* Investigating the viral ecology of global bee communities with high-throughput metagenomics. *Sci Rep* **8**, 8879, doi:10.1038/s41598-018-27164-z (2018).
56 de Miranda, J. R. *et al.* Standard methods for virus research in Apis mellifera. *Journal of Apicultural Research* **52**, 1-56, doi:10.3896/IBRA.1.52.4.22 (2013).
57 Wilfert, L. *et al.* Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites. *Science* **351**, 594-597, doi:10.1126/science.aac9976 (2016).
58 Hunter, W. *et al.* Large-scale field application of RNAi technology reducing Israeli acute paralysis virus disease in honey bees (Apis mellifera, Hymenoptera: Apidae). *PLoS Pathog* **6**, e1001160, doi:10.1371/journal.ppat.1001160 (2010).
59 Evans, J. D. *et al.* Immune pathways and defence mechanisms in honey bees Apis mellifera. *Insect Mol Biol* **15**, 645-656, doi:10.1111/j.1365-2583.2006.00682.x (2006).
60 Evans, J. D. & Spivak, M. Socialized medicine: individual and communal disease barriers in honey bees. *J Invertebr Pathol* **103 Suppl 1**, S62-72, doi:10.1016/j.jip.2009.06.019 (2010).
61 Danihlík, J., Aronstein, K. & Petřivalský, M. Antimicrobial peptides: a key component of honey bee innate immunity. *Journal of Apicultural Research* **54**, 123-136, doi:10.1080/00218839.2015.1109919 (2015).
62 Brutscher, L. M. & Flenniken, M. L. RNAi and Antiviral Defense in the Honey Bee. *J Immunol Res* **2015**, 941897, doi:10.1155/2015/941897 (2015).
63 Zufelato, M. S., Lourenço, A. P., Simões, Z. L., Jorge, J. A. & Bitondi, M. M. Phenoloxidase activity in Apis mellifera honey bee pupae, and ecdysteroid-dependent expression of the prophenoloxidase mRNA. *Insect Biochem Mol Biol* **34**, 1257-1268, doi:10.1016/j.ibmb.2004.08.005 (2004).
64 Wilson, E., Durlach, N. & Roth, L. Chemical releasers of necrophoric behavior in ants. *Psyche* **65**, 108-114 (1958).
65 Haskins, C. P. & Haskins, E. F. Notes on necrophoric behavior in the archaic ant Myrmecia vindex (Formicidae: Myrmeciinae). *Psyche* **81**, 258-267 (1974).
66 Gordon, D. M. Dependence of necrophoric response to oleic acid on social context in the ant, Pogonomyrmex badius. *Journal of chemical ecology* **9**, 105-111 (1983).
67 Howard, D. F. & Tschinkel, W. R. Aspects of necrophoric behavior in the red imported fire ant, Solenopsis invicta. *Behaviour* **56**, 157-178 (1976).
68 Julian, G. E. & Cahan, S. Undertaking specialization in the desert leaf-cutter ant Acromyrmex versicolor. *Animal Behaviour* **58**, 437-442 (1999).
69 Chouvenc, T., Robert, A., Sémon, E. & Bordereau, C. Burial behaviour by dealates of the termite Pseudacanthotermes spiniger (Termitidae, Macrotermitinae) induced by chemical signals from termite corpses. *Insectes sociaux* **59**, 119-125 (2012).
70 Ulyshen, M. D. & Shelton, T. G. Evidence of cue synergism in termite corpse response behavior. *Naturwissenschaften* **99**, 89-93 (2012).
71 Sun, Q., Haynes, K. F. & Zhou, X. Differential undertaking response of a lower termite to congeneric and conspecific corpses. *Sci Rep* **3**, 1650, doi:10.1038/srep01650 (2013).

72 Le Conte, Y. *et al.* Social immunity in honeybees (Apis mellifera): transcriptome analysis of varroa-hygienic behaviour. *Insect Mol Biol* **20**, 399-408, doi:10.1111/j.1365-2583.2011.01074.x (2011).
73 Rothenbuhler, W. C. Behavior genetics of nest cleaning in honey bees. IV. Responses of F 1 and backcross generations to disease-killed brood. *American Zoologist* **4**, 111-123 (1964).
74 Boecking, O. & Drescher, W. The removal response ofApis mellifera L. colonies to brood in wax and plastic cells after artificial and natural infestation withVarroa jacobsoni Oud. and to freeze-killed brood. *Experimental and Applied Acarology* **16**, 321-329 (1992).
75 Cremer, S., Armitage, S. A. & Schmid-Hempel, P. Social immunity. *Curr Biol* **17**, R693-702, doi:10.1016/j.cub.2007.06.008 (2007).
76 Meunier, J. Social immunity and the evolution of group living in insects. *Philos Trans R Soc Lond B Biol Sci* **370**, doi:10.1098/rstb.2014.0102 (2015).
77 Wilson-Rich, N., Spivak, M., Fefferman, N. H. & Starks, P. T. Genetic, individual, and group facilitation of disease resistance in insect societies. *Annu Rev Entomol* **54**, 405-423, doi:10.1146/annurev.ento.53.103106.093301 (2009).
78 Sun, Q. & Zhou, X. Corpse management in social insects. *Int J Biol Sci* **9**, 313-321, doi:10.7150/ijbs.5781 (2013).
79 Borba, R. S. & Spivak, M. Propolis envelope in Apis mellifera colonies supports honey bees against the pathogen, Paenibacillus larvae. *Sci Rep* **7**, 11429, doi:10.1038/s41598-017-11689-w (2017).
80 Simone, M., Evans, J. D. & Spivak, M. Resin collection and social immunity in honey bees. *Evolution* **63**, 3016-3022, doi:10.1111/j.1558-5646.2009.00772.x (2009).
81 Simone-Finstrom, M. D. & Spivak, M. Increased resin collection after parasite challenge: a case of self-medication in honey bees? *PLoS One* **7**, e34601, doi:10.1371/journal.pone.0034601 (2012).
82 Starks, P. T., Blackie, C. A. & Seeley, T. D. Fever in honeybee colonies. *Naturwissenschaften* **87**, 229-231 (2000).
83 Rothenbuhler, W. C. Behaviour genetics of nest cleaning in honey bees. I. Responses of four inbred lines to disease-killed brood. *Animal Behaviour* **12**, 578-583 (1964).
84 Villa, J. D., Danka, R. G. & Harris, J. W. Simplified methods of evaluating colonies for levels of Varroa Sensitive Hygiene (VSH). *Journal of apicultural research* **48**, 162-167 (2009).
85 Ibrahim, A. & Spivak, M. The relationship between hygienic behavior and suppression of mite reproduction as honey bee (Apis mellifera) mechanisms of resistance to Varroa destructor. *Apidologie* **37**, 31-40 (2006).
86 Stanimirović, Z., Stevanović, J., Mirilović, M. & Stojić, V. Heritability of hygienic behavior in grey honey bees (Apis mellifera carnica). *Acta veterinaria* **58**, 593-601 (2008).
87 Spivak, M. Honey bee hygienic behavior and defense against Varroa jacobsoni. *Apidologie* **27**, 245-260 (1996).
88 Spivak, M. & Reuter, G. S. Performance of hygienic honey bee colonies in a commercial apiary. *Apidologie* **29**, 291-302 (1998).

89 Harbo, J. R. & Harris, J. W. Suppressed mite reproduction explained by the behaviour of adult bees. *Journal of Apicultural Research* **44**, 21-23, doi:10.1080/00218839.2005.11101141 (2005).
90 Arathi, H. & Spivak, M. Influence of colony genotypic composition on the performance of hygienic behaviour in the honeybee, Apis mellifera L. *Animal Behaviour* **62**, 57-66 (2001).
91 Gramacho, K. P. & Spivak, M. Differences in olfactory sensitivity and behavioral responses among honey bees bred for hygienic behavior. *Behavioral Ecology and Sociobiology* **54**, 472-479 (2003).
92 Masterman, R., Ross, R., Mesce, K. & Spivak, M. Olfactory and behavioral response thresholds to odors of diseased brood differ between hygienic and non-hygienic honey bees (Apis mellifera L.). *Journal of Comparative Physiology A* **187**, 441-452 (2001).
93 Spivak, M., Masterman, R., Ross, R. & Mesce, K. A. Hygienic behavior in the honey bee (Apis mellifera L.) and the modulatory role of octopamine. *Journal of neurobiology* **55**, 341-354 (2003).
94 Swanson, J. A. *et al.* Odorants that induce hygienic behavior in honeybees: identification of volatile compounds in chalkbrood-infected honeybee larvae. *Journal of chemical ecology* **35**, 1108-1116 (2009).
95 Nazzi, F., Della Vedova, G. & D'Agaro, M. A semiochemical from brood cells infested by Varroa destructor triggers hygienic behaviour in Apis mellifera. *Apidologie* **35**, 65-70 (2004).
96 Mondet, F. *et al.* Specific Cues Associated With Honey Bee Social Defence against Varroa destructor Infested Brood. *Scientific reports* **6** (2016).
97 Yao, M. *et al.* The Ancient Chemistry of Avoiding Risks of Predation and Disease. *Evolutionary Biology* **36**, 267-281, doi:10.1007/s11692-009-9069-4 (2009).
98 Aksenov, V. & David Rollo, C. Necromone Death Cues and Risk Avoidance by the Cricket Acheta domesticus: Effects of Sex and Duration of Exposure. *Journal of Insect Behavior* **30**, 259-272, doi:10.1007/s10905-017-9612-6 (2017).
99 Qiu, H. L. *et al.* Differential necrophoric behaviour of the ant Solenopsis invicta towards fungal-infected corpses of workers and pupae. *Bull Entomol Res* **105**, 607-614, doi:10.1017/S0007485315000528 (2015).
100 Akino, T. & Yamaoka, R. Origin of Oleic Acid, Corpse Recognition Signal in the Ant, Formica japonica MOTSCHLSKY (Hymenoptera: Formicidae). *Japanese journal of applied entomology and zoology* **40**, 265-271, doi:10.1303/jjaez.40.265 (1996).
101 Abbott, K. R. Bumblebees avoid flowers containing evidence of past predation events. *Canadian Journal of Zoology* **84**, 1240-1247, doi:10.1139/z06-117 (2006).
102 Rollo, C. D., Czvzewska, E. & Borden, J. H. Fatty acid necromones for cockroaches. *Naturwissenschaften* **81**, 409-410, doi:10.1007/bf01132695 (1994).
103 Spivak, M. & Reuter, G. S. New direction for the Minnesota hygienic line of bees.
104 Danka, R. G., Harris, J. W. & Villa, J. D. Expression of Varroa sensitive hygiene (VSH) in commercial VSH honey bees (Hymenoptera: Apidae). *Journal of economic entomology* **104**, 745-749 (2011).
105 Rinderer, T. E., Harris, J. W., Hunt, G. J. & De Guzman, L. I. Breeding for resistance to Varroa destructor in North America. *Apidologie* **41**, 409-424 (2010).

106 Guarna, M. M. *et al.* A search for protein biomarkers links olfactory signal transduction to social immunity. *BMC Genomics* **16**, 63, doi:10.1186/s12864-014-1193-6 (2015).
107 Behrens, D. & Moritz, R. F. A. QTL-mapping of individual resistance against American foulbrood in haploid honeybee drone larvae (Apis mellifera). *Apidologie* **45**, 409-417, doi:10.1007/s13592-013-0255-0 (2014).
108 Tsuruda, J. M., Harris, J. W., Bourgeois, L., Danka, R. G. & Hunt, G. J. High-resolution linkage analyses to identify genes that influence Varroa sensitive hygiene behavior in honey bees. *PLoS One* **7**, e48276, doi:10.1371/journal.pone.0048276 (2012).
109 Behrens, D. *et al.* Three QTL in the honey bee Apis mellifera L. suppress reproduction of the parasitic mite Varroa destructor. *Ecol Evol* **1**, 451-458, doi:10.1002/ece3.17 (2011).
110 Beye, M. *et al.* Exceptionally high levels of recombination across the honey bee genome. *Genome Res* **16**, 1339-1344, doi:10.1101/gr.5680406 (2006).
111 Gempe, T., Stach, S., Bienefeld, K., Otte, M. & Beye, M. Behavioral and molecular studies of quantitative differences in hygienic behavior in honeybees. *BMC Res Notes* **9**, 474, doi:10.1186/s13104-016-2269-y (2016).
112 Guarna, M. M. *et al.* Peptide biomarkers used for the selective breeding of a complex polygenic trait in honey bees. *Sci Rep* **7**, 8381, doi:10.1038/s41598-017-08464-2 (2017).
113 Hu, H. *et al.* Proteome Analysis of the Hemolymph, Mushroom Body, and Antenna Provides Novel Insight into Honeybee Resistance against Varroa Infestation. *J Proteome Res* **15**, 2841-2854, doi:10.1021/acs.jproteome.6b00423 (2016).
114 Ji, T. *et al.* Transcriptional responses in eastern honeybees (Apis cerana) infected with mites, Varroa destructor. *Genet Mol Res* **13**, 8888-8900, doi:10.4238/2014.October.31.4 (2014).
115 Hu, H. *et al.* Proteome analysis of the hemolymph, mushroom body, and antenna provides novel insight into honeybee resistance against Varroa infestation. *Journal of Proteome Research* (2016).
116 Winston, M. L. *The biology of the honey bee*. (harvard university press, 1991).
117 Robertson, H. M. & Wanner, K. W. The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family. *Genome Res* **16**, 1395-1403, doi:10.1101/gr.5057506 (2006).
118 Frasnelli, E. *et al.* The Bee as a Model to Investigate Brain and Behavioural Asymmetries. *Insects* **5**, 120-138, doi:10.3390/insects5010120 (2014).
119 Abel, R., Rybak, J. & Menzel, R. Structure and response patterns of olfactory interneurons in the honeybee, Apis mellifera. *J Comp Neurol* **437**, 363-383 (2001).
120 Forêt, S. & Maleszka, R. Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (Apis mellifera). *Genome Res* **16**, 1404-1413, doi:10.1101/gr.5075706 (2006).
121 Fan, J., Francis, F., Liu, Y., Chen, J. L. & Cheng, D. F. An overview of odorant-binding protein functions in insect peripheral olfactory reception. *Genet Mol Res* **10**, 3056-3069, doi:10.4238/2011.December.8.2 (2011).
122 Pelosi, P., Calvello, M. & Ban, L. Diversity of odorant-binding proteins and chemosensory proteins in insects. *Chem Senses* **30 Suppl 1**, i291-292, doi:10.1093/chemse/bjh229 (2005).

123 Brito, N. F., Moreira, M. F. & Melo, A. C. A look inside odorant-binding proteins in insect chemoreception. *J Insect Physiol* **95**, 51-65, doi:10.1016/j.jinsphys.2016.09.008 (2016).
124 Vogt, R. G. Molecular basis of pheromone detection in insects.
125 Chan, Q. W. *et al.* Honey bee protein atlas at organ-level resolution. *Genome Res* **23**, 1951-1960, doi:10.1101/gr.155994.113 (2013).
126 Han, L. *et al.* Operating mechanism and molecular dynamics of pheromone-binding protein ASP1 as influenced by pH. *PLoS One* **9**, e110565, doi:10.1371/journal.pone.0110565 (2014).
127 Cox-Foster, D. L. *et al.* A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**, 283-287, doi:10.1126/science.1146498 (2007).
128 Evans, J. D. & Schwarz, R. S. Bees brought to their knees: microbes affecting honey bee health. *Trends Microbiol* **19**, 614-620, doi:10.1016/j.tim.2011.09.003 (2011).
129 Tsvetkov, N. *et al.* Chronic exposure to neonicotinoids reduces honey bee health near corn crops. *Science* **356**, 1395-1397, doi:10.1126/science.aam7470 (2017).
130 Woodcock, B. A. *et al.* Country-specific effects of neonicotinoid pesticides on honey bees and wild bees. *Science* **356**, 1393-1395, doi:10.1126/science.aaa1190 (2017).
131 Johnson, R. M., Ellis, M. D., Mullin, C. A. & Frazier, M. Pesticides and honey bee toxicity–USA. *Apidologie* **41**, 312-331 (2010).
132 Blacquiere, T., Smagghe, G., Van Gestel, C. A. & Mommaerts, V. Neonicotinoids in bees: a review on concentrations, side-effects and risk assessment. *Ecotoxicology* **21**, 973-992 (2012).
133 Connolly, C. N. The risk of insecticides to pollinating insects. *Commun Integr Biol* **6**, e25074, doi:10.4161/cib.25074 (2013).
134 Adams, M. D. *et al.* The genome sequence of Drosophila melanogaster. *Science* **287**, 2185-2195 (2000).
135 Holt, R. A. *et al.* The genome sequence of the malaria mosquito Anopheles gambiae. *Science* **298**, 129-149, doi:10.1126/science.1076181 (2002).
136 Xia, Q. *et al.* A draft sequence for the genome of the domesticated silkworm (Bombyx mori). *Science* **306**, 1937-1940, doi:10.1126/science.1102210 (2004).
137 Brent, M. R. How does eukaryotic gene prediction work? *Nat Biotechnol* **25**, 883-885, doi:10.1038/nbt0807-883 (2007).
138 Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59, doi:10.1186/1471-2105-5-59 (2004).
139 Zickmann, F. & Renard, B. Y. IPred - integrating ab initio and evidence based gene predictions to improve prediction accuracy. *BMC Genomics* **16**, 134, doi:10.1186/s12864-015-1315-9 (2015).
140 Jørgensen, F. G., Schierup, M. H. & Clark, A. G. Heterogeneity in regional GC content and differential usage of codons and amino acids in GC-poor and GC-rich regions of the genome of Apis mellifera. *Mol Biol Evol* **24**, 611-619, doi:10.1093/molbev/msl190 (2007).
141 Brown, J. B. *et al.* Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**, 393-399, doi:10.1038/nature12962 (2014).
142 Trapp, J. *et al.* High-throughput proteome dynamics for discovery of key proteins in sentinel species: Unsuspected vitellogenins diversity in the crustacean Gammarus

fossarum. *Journal of Proteomics* **146**, 207-214, doi:http://dx.doi.org/10.1016/j.jprot.2016.07.007 (2016).
143 Chan, Q. W., Howes, C. G. & Foster, L. J. Quantitative comparison of caste differences in honeybee hemolymph. *Mol Cell Proteomics* **5**, 2252-2262, doi:10.1074/mcp.M600197-MCP200 (2006).
144 Hummon, A. B. *et al.* From the genome to the proteome: uncovering peptides in the Apis brain. *Science* **314**, 647-649, doi:10.1126/science.1124128 (2006).
145 Wolschin, F. & Amdam, G. V. Comparative proteomics reveal characteristics of life-history transitions in a social insect. *Proteome Sci* **5**, 10, doi:10.1186/1477-5956-5-10 (2007).
146 Chan, Q. W. *et al.* The worker honeybee fat body proteome is extensively remodeled preceding a major life-history transition. *PLoS One* **6**, e24794, doi:10.1371/journal.pone.0024794 (2011).
147 Wolschin, F. & Amdam, G. V. Plasticity and robustness of protein patterns during reversible development in the honey bee (Apis mellifera). *Anal Bioanal Chem* **389**, 1095-1100, doi:10.1007/s00216-007-1523-5 (2007).
148 Ament, S. A., Corona, M., Pollock, H. S. & Robinson, G. E. Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc Natl Acad Sci U S A* **105**, 4226-4231, doi:10.1073/pnas.0800630105 (2008).
149 Wolschin, F., Münch, D. & Amdam, G. V. Structural and proteomic analyses reveal regional brain differences during honeybee aging. *J Exp Biol* **212**, 4027-4032, doi:10.1242/jeb.033845 (2009).
150 Hu, H. *et al.* Proteome Analysis of the Hemolymph, Mushroom Body, and Antenna Provides Novel Insight into Honeybee Resistance against Varroa Infestation. *J Proteome Res*, doi:10.1021/acs.jproteome.6b00423 (2016).
151 McAfee, A. *et al.* Toward an Upgraded Honey Bee (Apis mellifera L.) Genome Annotation Using Proteogenomics. *J Proteome Res* **15**, 411-421, doi:10.1021/acs.jproteome.5b00589 (2016).
152 Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587, doi:10.1038/nature13319 (2014).
153 Kim, M. S. *et al.* A draft map of the human proteome. *Nature* **509**, 575-581, doi:10.1038/nature13302 (2014).
154 Merrihew, G. E. *et al.* Use of shotgun proteomics for the identification, confirmation, and correction of C. elegans gene annotations. *Genome Res* **18**, 1660-1669, doi:10.1101/gr.077644.108 (2008).
155 Brunner, E. *et al.* A high-quality catalog of the Drosophila melanogaster proteome. *Nat Biotechnol* **25**, 576-583, doi:10.1038/nbt1300 (2007).
156 Castellana, N. E. *et al.* Discovery and revision of Arabidopsis genes by proteogenomics. *Proc Natl Acad Sci U S A* **105**, 21034-21038, doi:10.1073/pnas.0811066106 (2008).
157 Brosch, M. *et al.* Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and "resurrected" pseudogenes in the mouse genome. *Genome Res* **21**, 756-767, doi:10.1101/gr.114272.110 (2011).
158 Deutsch, E. W. *et al.* Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. *J Proteome Res* **15**, 3961-3970, doi:10.1021/acs.jproteome.6b00392 (2016).

159 Jaffe, J. D., Berg, H. C. & Church, G. M. Proteogenomic mapping as a complementary method to perform genome annotation. *PROTEOMICS* **4**, 59-77 (2004).
160 Armengaud, J. A perfect genome annotation is within reach with the proteomics and genomics alliance. *Current Opinion in Microbiology* **12**, 292-300 (2009).
161 Yates, J. R., Eng, J. K. & McCormack, A. L. Mining Genomes: Correlating Tandem Mass Spectra of Modified and Unmodified Peptides to Sequences in Nucleotide Databases. *Analytical Chemistry* **67**, 3202-3210, doi:10.1021/ac00114a016 (1995).
162 Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* **13**, 397-406, doi:10.1074/mcp.M113.035600 (2014).
163 Hestand, M. S., Zeng, Z., Coleman, S. J., Liu, J. & MacLeod, J. N. Tissue Restricted Splice Junctions Originate Not Only from Tissue-Specific Gene Loci, but Gene Loci with a Broad Pattern of Expression. *PLoS One* **10**, e0144302, doi:10.1371/journal.pone.0144302 (2015).
164 Chan, Q. W., Parker, R., Sun, Z., Deutsch, E. W. & Foster, L. J. A honey bee (Apis mellifera L.) PeptideAtlas crossing castes and tissues. *BMC Genomics* **12**, 290, doi:10.1186/1471-2164-12-290 (2011).
165 Garbian, Y., Maori, E., Kalev, H., Shafir, S. & Sela, I. Bidirectional transfer of RNAi between honey bee and Varroa destructor: Varroa gene silencing reduces Varroa population. *PLoS Pathog* **8**, e1003035, doi:10.1371/journal.ppat.1003035 (2012).
166 Campbell, E. M., Budge, G. E. & Bowman, A. S. Gene-knockdown in the honey bee mite Varroa destructor by a non-invasive approach: studies on a glutathione S-transferase. *Parasit Vectors* **3**, 73, doi:10.1186/1756-3305-3-73 (2010).
167 Campbell, E. M., Budge, G. E., Watkins, M. & Bowman, A. S. Transcriptome analysis of the synganglion from the honey bee mite, Varroa destructor and RNAi knockdown of neural peptide targets. *Insect Biochem Mol Biol* **70**, 116-126, doi:10.1016/j.ibmb.2015.12.007 (2016).
168 Ziegelmann, B. *et al.* Lithium chloride effectively kills the honey bee parasite Varroa destructor by a systemic mode of action. *Sci Rep* **8**, 683, doi:10.1038/s41598-017-19137-5 (2018).
169 Levin, S., Sela, N. & Chejanovsky, N. Two novel viruses associated with the Apis mellifera pathogenic mite Varroa destructor. *Sci Rep* **6**, 37710, doi:10.1038/srep37710 (2016).
170 Erban, T. *et al.* In-depth proteomic analysis of Varroa destructor: Detection of DWV-complex, ABPV, VdMLV and honeybee proteins in the mite. *Sci Rep* **5**, 13907, doi:10.1038/srep13907 (2015).
171 Gisder, S., Aumeier, P. & Genersch, E. Deformed wing virus: replication and viral load in mites (Varroa destructor). *J Gen Virol* **90**, 463-467, doi:10.1099/vir.0.005579-0 (2009).
172 Ryabov, E. V. *et al.* A virulent strain of deformed wing virus (DWV) of honeybees (Apis mellifera) prevails after Varroa destructor-mediated, or in vitro, transmission. *PLoS Pathog* **10**, e1004230, doi:10.1371/journal.ppat.1004230 (2014).
173 Ziegelmann, B., Lindenmayer, A., Steidle, J. & Rosenkranz, P. The mating behavior of Varroa destructor is triggered by a female sex pheromone. *Apidologie* **44**, 314-323, doi:10.1007/s13592-012-0182-5 (2013).

174 Del Piccolo, F., Nazzi, F., Della Vedova, G. & Milani, N. Selection of Apis mellifera workers by the parasitic mite Varroa destructor using host cuticular hydrocarbons. *Parasitology* **137**, 967-973, doi:10.1017/S0031182009991867 (2010).
175 Nazzi, F. & Le Conte, Y. Ecology of Varroa destructor, the Major Ectoparasite of the Western Honey Bee, Apis mellifera. *Annu Rev Entomol* **61**, 417-432, doi:10.1146/annurev-ento-010715-023731 (2016).
176 Andino, G. K., Gribskov, M., Anderson, D. L., Evans, J. D. & Hunt, G. J. Differential gene expression in Varroa jacobsoni mites following a host shift to European honey bees (Apis mellifera). *BMC Genomics* **17**, 926, doi:10.1186/s12864-016-3130-3 (2016).
177 Surlis, C., Carolan, J. C., Coffey, M. F. & Kavanagh, K. Proteomic analysis of Bayvarol® resistance mechanisms in the honey bee parasite Varroa destructor. *Journal of Apicultural Research* **55**, 49-64 (2016).
178 McAfee, A. & Foster, L. J. Proteogenomics: Recycling Public Data to Improve Genome Annotations. *Methods Enzymol* **585**, 217-243, doi:10.1016/bs.mie.2016.09.020 (2017).
179 Nesvizhskii, A. I. Proteogenomics: concepts, applications and computational strategies. *Nat Methods* **11**, 1114-1125, doi:10.1038/nmeth.3144 (2014).
180 Fang, Y. *et al.* In-depth proteomics characterization of embryogenesis of the honey bee worker (Apis mellifera ligustica). *Mol Cell Proteomics* **13**, 2306-2320, doi:10.1074/mcp.M114.037846 (2014).
181 Geiger, T. *et al.* Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. *Mol Cell Proteomics* **12**, 1709-1722, doi:10.1074/mcp.M112.024919 (2013).
182 de Godoy, L. M. *et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251-1254, doi:10.1038/nature07341 (2008).
183 Kristensen, A. R., Gsponer, J. & Foster, L. J. Protein synthesis rate is the predominant regulator of protein expression during differentiation. *Mol Syst Biol* **9**, 689, doi:10.1038/msb.2013.47 (2013).
184 Hillier, L. W. *et al.* Genomics in C. elegans: so many genes, such a little worm. *Genome Res* **15**, 1651-1660, doi:10.1101/gr.3729105 (2005).
185 Church, D. M. *et al.* Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol* **7**, e1000112, doi:10.1371/journal.pbio.1000112 (2009).
186 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
187 Bogaerts, A., Baggerman, G., Vierstraete, E., Schoofs, L. & Verleyen, P. The hemolymph proteome of the honeybee: gel-based or gel-free? *Proteomics* **9**, 3201-3208, doi:10.1002/pmic.200800604 (2009).
188 Yandell, M. & Ence, D. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet* **13**, 329-342 (2012).
189 Ekblom, R. & Wolf, J. B. A field guide to whole-genome sequencing, assembly and annotation. *Evol Appl* **7**, 1026-1042, doi:10.1111/eva.12178 (2014).
190 Yue, L. *et al.* Genetic analysis of viable Hsp90 alleles reveals a critical role in Drosophila spermatogenesis. *Genetics* **151**, 1065-1079 (1999).
191 Dynes, T. L. *et al.* Fine scale population genetic structure of Varroa destructor, an ectoparasitic mite of the honey bee (Apis mellifera). *Apidologie* **2016**, 1-9 (2016).

192 McAfee, A., Collins, T. F., Madilao, L. L. & Foster, L. J. Odorant cues linked to social immunity induce lateralized antenna stimulation in honey bees (Apis mellifera L.). *bioRxiv*, doi:10.1101/086330 (2016).
193 Foster, L. J., De Hoog, C. L. & Mann, M. Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. *Proc Natl Acad Sci U S A* **100**, 5813-5818, doi:10.1073/pnas.0631608100 (2003).
194 Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-670 (2003).
195 Beck, S. *et al.* The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics. *Mol Cell Proteomics* **14**, 2014-2029, doi:10.1074/mcp.M114.047407 (2015).
196 Lee, H. K., Braynen, W., Keshav, K. & Pavlidis, P. ErmineJ: tool for functional analysis of gene expression data sets. *BMC Bioinformatics* **6**, 269, doi:10.1186/1471-2105-6-269 (2005).
197 Michaud, S. *et al.* Response of the honey bee (Apis mellifera) proteome to Israeli acute paralysis virus (IAPV) infection. *Canadian Journal of Zoology* **93**, 711-720, doi:10.1139/cjz-2014-0181 (2014).
198 Cox-Foster, D. L. *et al.* A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**, 283-287 (2007).
199 Guarna, M. M. *et al.* A search for protein biomarkers links olfactory signal transduction to social immunity. *BMC genomics* **16**, 1 (2015).
200 Chakroborty, N. K., Bienefeld, K. & Menzel, R. Odor learning and odor discrimination of bees selected for enhanced hygienic behavior. *Apidologie* **46**, 499-514 (2015).
201 Sandoz, J. C. Behavioral and neurophysiological study of olfactory perception and learning in honeybees. *Front Syst Neurosci* **5**, 98, doi:10.3389/fnsys.2011.00098 (2011).
202 Parker, R. *et al.* Correlation of proteome-wide changes with social immunity behaviors provides insight into resistance to the parasitic mite, Varroa destructor, in the honey bee (Apis mellifera). *Genome biology* **13**, 1 (2012).
203 Martin, C. *et al.* Potential mechanism for detection by Apis mellifera of the parasitic mite Varroa destructor inside sealed brood cells. *Physiological Entomology* **27**, 175-188 (2002).
204 Annoscia, D., Del Piccolo, F. & Nazzi, F. How does the mite Varroa destructor kill the honeybee Apis mellifera? Alteration of cuticular hydrcarbons and water loss in infested honeybees. *Journal of insect physiology* **58**, 1548-1555 (2012).
205 Salvy, M. *et al.* Modifications of the cuticular hydrocarbon profile of Apis mellifera worker bees in the presence of the ectoparasitic mite Varroa jacobsoni in brood cells. *Parasitology* **122**, 145-159 (2001).
206 He, X. J. *et al.* Starving honey bee (Apis mellifera) larvae signal pheromonally to worker bees. *Scientific reports* **6** (2016).
207 Frasnelli, E., Anfora, G., Trona, F., Tessarolo, F. & Vallortigara, G. Morpho-functional asymmetry of the olfactory receptors of the honeybee (Apis mellifera). *Behav Brain Res* **209**, 221-225, doi:10.1016/j.bbr.2010.01.046 (2010).
208 Anfora, G., Frasnelli, E., Maccagnani, B., Rogers, L. J. & Vallortigara, G. Behavioural and electrophysiological lateralization in a social (Apis mellifera) but not in a non-social

(Osmia cornuta) species of bee. *Behavioural Brain Research* **206**, 236-239, doi:http://dx.doi.org/10.1016/j.bbr.2009.09.023 (2010).
209 Kundu, S., Ganguly, A., Chakraborty, T. S., Kumar, A. & Siddiqi, O. Synergism and Combinatorial Coding for Binary Odor Mixture Perception in Drosophila. *eNeuro* **3**, doi:10.1523/ENEURO.0056-14.2016 (2016).
210 Krofczik, S., Menzel, R. & Nawrot, M. P. Rapid odor processing in the honeybee antennal lobe network. *Front Comput Neurosci* **2**, 9, doi:10.3389/neuro.10.009.2008 (2008).
211 Reinhard, J., Sinclair, M., Srinivasan, M. V. & Claudianos, C. Honeybees learn odour mixtures via a selection of key odorants. *PLoS One* **5**, e9110, doi:10.1371/journal.pone.0009110 (2010).
212 Deisig, N., Giurfa, M. & Sandoz, J. C. Antennal lobe processing increases separability of odor mixture representations in the honeybee. *J Neurophysiol* **103**, 2185-2194, doi:10.1152/jn.00342.2009 (2010).
213 Anfora, G. *et al.* Lateralization in the invertebrate brain: left-right asymmetry of olfaction in bumble bee, Bombus terrestris. *PLoS One* **6**, e18903, doi:10.1371/journal.pone.0018903 (2011).
214 Frasnelli, E., Vallortigara, G. & Rogers, L. J. Origins of brain asymmetry: lateralization of odour memory recall in primitive Australian stingless bees. *Behav Brain Res* **224**, 121-127, doi:10.1016/j.bbr.2011.05.026 (2011).
215 Letzkus, P. *et al.* Lateralization of olfaction in the honeybee Apis mellifera. *Curr Biol* **16**, 1471-1476, doi:10.1016/j.cub.2006.05.060 (2006).
216 Rogers, L. J., Frasnelli, E. & Versace, E. Lateralized antennal control of aggression and sex differences in red mason bees, Osmia bicornis. *Sci Rep* **6**, 29411, doi:10.1038/srep29411 (2016).
217 Rogers, L. J. & Vallortigara, G. From antenna to antenna: lateral shift of olfactory memory recall by honeybees. *PLoS One* **3**, e2340, doi:10.1371/journal.pone.0002340 (2008).
218 Rogers, L. J., Rigosi, E., Frasnelli, E. & Vallortigara, G. A right antenna for social behaviour in honeybees. *Sci Rep* **3**, 2045, doi:10.1038/srep02045 (2013).
219 Meng, Z. J., Yan, S. C., Yang, C. P. & Ruan, C. C. Asymmetrical distribution of antennal sensilla in the female Anastatus japonicus Ashmead (Hymenoptera: Eupelmidae). *Microsc Res Tech* **75**, 1066-1075, doi:10.1002/jemt.22032 (2012).
220 Carroll, M. J. & Duehl, A. J. Collection of volatiles from honeybee larvae and adults enclosed on brood frames. *Apidologie* **43**, 715-730 (2012).
221 Maisonnasse, A., Lenoir, J.-C., Beslay, D., Crauser, D. & Le Conte, Y. E-β-ocimene, a volatile brood pheromone involved in social regulation in the honey bee colony (Apis mellifera). *PLoS One* **5**, e13531 (2010).
222 Maisonnasse, A. *et al.* A scientific note on E-$\ beta $-ocimene, a new volatile primer pheromone that inhibits worker ovary development in honey bees. *Apidologie* **40**, 562-564 (2009).
223 Trhlin, M. & Rajchard, J. Chemical communication in the honeybee (Apis mellifera L.): a review. *Vet. Med* **56**, 265-273 (2011).

224 Traynor, K. S., Le Conte, Y. & Page, R. E. Age matters: pheromone profiles of larvae differentially influence foraging behaviour in the honeybee, Apis mellifera. *Animal Behaviour* **99**, 1-8 (2015).
225 Le Conte, Y. *et al.* Varroa destructor changes its cuticular hydrocarbons to mimic new hosts. *Biol Lett* **11**, 20150233, doi:10.1098/rsbl.2015.0233 (2015).
226 Panasiuk, B., Skowronek, W., Bienkowska, M., Gerula, D. & Wegrzynovicz, P. Age of worker bees performing hygienic behaviour in a honeybee colony. *Journal of Apicultural Science* **54**, 109-115 (2010).
227 Olsson, S. B. & Hansson, B. S. Electroantennogram and single sensillum recording in insect antennae. *Methods Mol Biol* **1068**, 157-177, doi:10.1007/978-1-62703-619-1_11 (2013).
228 Humphrey, S. J., Azimifar, S. B. & Mann, M. High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nat Biotechnol* **33**, 990-995, doi:10.1038/nbt.3327 (2015).
229 Yang, F., Shen, Y., Camp, D. G. & Smith, R. D. High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Rev Proteomics* **9**, 129-134, doi:10.1586/epr.12.15 (2012).
230 McAfee, A., Collins, T. F., Madilao, L. L. & Foster, L. J. Odorant cues linked to social immunity induce lateralized antenna stimulation in honey bees (Apis mellifera L.). *Sci Rep* **7**, 46171, doi:10.1038/srep46171 (2017).
231 Boecking, O. & Drescher, W. Response of Apis mellifera L colonies infested with Varroa jacobsoni Oud. *Apidologie* **22**, 237-241 (1991).
232 Gilley, D. C., Degrandi-Hoffman, G. & Hooper, J. E. Volatile compounds emitted by live European honey bee (Apis mellifera L.) queens. *J Insect Physiol* **52**, 520-527, doi:10.1016/j.jinsphys.2006.01.014 (2006).
233 He, X. J. *et al.* Starving honey bee (Apis mellifera) larvae signal pheromonally to worker bees. *Sci Rep* **6**, 22359, doi:10.1038/srep22359 (2016).
234 Ma, R., Mueller, U. G. & Rangel, J. Assessing the role of β-ocimene in regulating foraging behavior of the honey bee, Apis mellifera. *Apidologie* **47**, 135-144, doi:10.1007/s13592-015-0382-x (2016).
235 Maisonnasse, A., Lenoir, J. C., Beslay, D., Crauser, D. & Le Conte, Y. E-β-ocimene, a volatile brood pheromone involved in social regulation in the honey bee colony (Apis mellifera). *PLoS One* **5**, e13531, doi:10.1371/journal.pone.0013531 (2010).
236 Plettner, E., Eliash, N., Singh, N. K., Pinnelli, G. R. & Soroker, V. The chemical ecology of host-parasite interaction as a target of Varroa destructor control agents. *Apidologie* **48**, 78-92 (2017).
237 Zalewski, K., Zaobidna, E. & Żółtowska, K. Fatty acid composition of the parasitic mite Varroa destructor and its host the worker prepupae of Apis mellifera. *Physiological Entomology* **41**, 31-37 (2016).
238 Dmitryjuk, M., Zalewski, K., Raczkowski, M. & Zoltowska, K. Composition of fatty acids in the Varroa destructor mites and their hosts, Apis mellifera drone-prepupae. *Annals of parasitology* **61** (2015).
239 Jones, W. Olfactory carbon dioxide detection by insects and other animals. *Mol Cells* **35**, 87-92, doi:10.1007/s10059-013-0035-8 (2013).

240 Ament, S. A. *et al.* The transcription factor ultraspiracle influences honey bee social behavior and behavior-related gene expression. *PLoS Genet* **8**, e1002596, doi:10.1371/journal.pgen.1002596 (2012).
241 Nelson, C. M., Ihle, K. E., Fondrk, M. K., Page, R. E. & Amdam, G. V. The gene vitellogenin has multiple coordinating effects on social organization. *PLoS Biol* **5**, e62, doi:10.1371/journal.pbio.0050062 (2007).
242 Li-Byarlay, H. *et al.* RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proc Natl Acad Sci U S A* **110**, 12750-12755, doi:10.1073/pnas.1310735110 (2013).
243 Amdam, G. V., Simões, Z. L., Guidugli, K. R., Norberg, K. & Omholt, S. W. Disruption of vitellogenin gene function in adult honeybees by intra-abdominal injection of double-stranded RNA. *BMC Biotechnol* **3**, 1 (2003).
244 Dearden, P. K., Duncan, E. J. & Wilson, M. J. RNA interference (RNAi) in honeybee (Apis mellifera) embryos. *Cold Spring Harb Protoc* **2009**, pdb.prot5228, doi:10.1101/pdb.prot5228 (2009).
245 Maleszka, J., Forêt, S., Saint, R. & Maleszka, R. RNAi-induced phenotypes suggest a novel role for a chemosensory protein CSP5 in the development of embryonic integument in the honeybee (Apis mellifera). *Dev Genes Evol* **217**, 189-196, doi:10.1007/s00427-006-0127-y (2007).
246 Li, W. *et al.* Silencing the Honey Bee (Apis mellifera) Naked Cuticle Gene (nkd) Improves Host Immune Function and Reduces Nosema ceranae Infections. *Appl Environ Microbiol* **82**, 6779-6787, doi:10.1128/AEM.02105-16 (2016).
247 Gatehouse, H. S. *et al.* Amylase activity in honey bee hypopharyngeal glands reduced by RNA interference. *Journal of Apicultural Research* **43**, 9-13, doi:10.1080/00218839.2004.11101101 (2004).
248 Rein, J., Mustard, J. A., Strauch, M., Smith, B. H. & Galizia, C. G. Octopamine modulates activity of neural networks in the honey bee antennal lobe. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **199**, 947-962, doi:10.1007/s00359-013-0805-y (2013).
249 Huvenne, H. & Smagghe, G. Mechanisms of dsRNA uptake in insects and potential of RNAi for pest control: a review. *J Insect Physiol* **56**, 227-235, doi:10.1016/j.jinsphys.2009.10.004 (2010).
250 Schulte, C., Theilenberg, E., Müller-Borg, M., Gempe, T. & Beye, M. Highly efficient integration and expression of piggyBac-derived cassettes in the honeybee (Apis mellifera). *Proc Natl Acad Sci U S A* **111**, 9003-9008, doi:10.1073/pnas.1402341111 (2014).
251 Schulte, C. *et al.* Honey bee promoter sequences for targeted gene expression. *Insect Mol Biol* **22**, 399-410, doi:10.1111/imb.12031 (2013).
252 Price, D. R. & Gatehouse, J. A. RNAi-mediated crop protection against insects. *Trends Biotechnol* **26**, 393-400, doi:10.1016/j.tibtech.2008.04.004 (2008).
253 Schmehl, D. R., Tomé, H. V. V., Mortensen, A. N., Martins, G. F. & Ellis, J. D. Protocol for the in vitro rearing of honey bee (Apis mellifera L.) workers. *Journal of Apicultural Research* **55**, 113-129, doi:10.1080/00218839.2016.1203530 (2016).

254 Kaftanoglu, O., Linksvayer, T. A. & Page, R. E. Rearing Honey Bees, Apis mellifera, in vitro 1: Effects of Sugar Concentrations on Survival and Development. *Journal of Insect Science* **11**, 96, doi:10.1673/031.011.9601 (2011).
255 Crailsheim, K. *et al.* Standard methods for artificial rearing of Apis mellifera larvae. *Journal of apicultural research* **52**, 1-16 (2013).
256 Iovinella, I., Bozza, F., Caputo, B., Della Torre, A. & Pelosi, P. Ligand-binding study of Anopheles gambiae chemosensory proteins. *Chem Senses* **38**, 409-419, doi:10.1093/chemse/bjt012 (2013).
257 Leal, G. M. & Leal, W. S. Binding of a fluorescence reporter and a ligand to an odorant-binding protein of the yellow fever mosquito, Aedes aegypti. *F1000Research* **3** (2014).
258 Sun, Y. F. *et al.* Two odorant-binding proteins mediate the behavioural response of aphids to the alarm pheromone (E)-ß-farnesene and structural analogues. *PloS one* **7**, e32759 (2012).
259 Wang, Y., Wang, F., Wang, R., Zhao, P. & Xia, Q. 2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori. *Sci Rep* **5**, 16273, doi:10.1038/srep16273 (2015).
260 Wang, Q. S. & Jan, E. Switch from cap-to factorless IRES-dependent 0 and+ 1 frame translation during cellular stress and dicistrovirus infection. *PloS one* **9**, e103601 (2014).
261 Kim, J. H. *et al.* High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* **6**, e18556, doi:10.1371/journal.pone.0018556 (2011).
262 Ban, L. *et al.* Chemosensory proteins of Locusta migratoria. *Insect Mol Biol* **12**, 125-134 (2003).
263 Zhou, R., Mohr, S., Hannon, G. J. & Perrimon, N. Inducing RNAi in Drosophila cells by soaking with dsRNA. *Cold Spring Harb Protoc* **2014**, doi:10.1101/pdb.prot080747 (2014).
264 Peterson, K. J. *et al.* Estimating metazoan divergence times with a molecular clock. *Proc Natl Acad Sci U S A* **101**, 6536-6541, doi:10.1073/pnas.0401670101 (2004).
265 Kapheim, K. M. *et al.* Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139-1143, doi:10.1126/science.aaa4788 (2015).
266 Leclercq, G., Pannebakker, B., Gengler, N., Nguyen, B. K. & Francis, F. Drawbacks and benefits of hygienic behavior in honey bees (Apis mellifera L.): a review. *Journal of Apicultural Research* **56**, 366-375 (2017).
267 Danka, R. G., Harris, J. W., Villa, J. D. & Dodds, G. E. Varying congruence of hygienic responses to Varroa destructor and freeze-killed brood among different types of honeybees. *Apidologie* **44**, 447-457 (2013).
268 Ruvolo-Takasusuki, M. C. C., Toledo, V. A. & Lopes, D. A. Relationship between hygienic behavior and Varroa destructor mites in colonies producing honey or royal jelly. *Sociobiology* **59**, 251-274 (2014).